

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 8.49515 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928A-1

Perfect score: 38

Sequence: 1 YEDEINK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80.\*

1: Pirl.\*

2: Pirl2.\*

3: Pirl3.\*

4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	310	2 S43865	cyto keratin 8, typ
2	38	100.0	384	2 I61769	keratin 6d, type I
3	38	100.0	390	2 A60093	cyto keratin, type
4	38	100.0	419	1 KRXL2B	keratin, 64K type
5	38	100.0	425	1 KRXL2A	keratin, 64K type
6	38	100.0	461	2 A43782	keratin, type II -
7	38	100.0	483	2 A34720	keratin 8, type II
8	38	100.0	487	2 JTO407	keratin 8, type II
9	38	100.0	490	2 JS0658	cyto keratin EndoA
10	38	100.0	513	2 S08381	keratin, 58K type
11	38	100.0	520	2 JS0291	intermediate filam
12	38	100.0	553	2 I59009	epidermal keratin
13	38	100.0	564	1 KRHUEA	keratin 6a, type I
14	38	100.0	564	1 KRHUEB	keratin 6b, type I
15	38	100.0	564	2 I61770	keratin 6e, type I
16	38	100.0	564	2 I61768	keratin 6c, type I
17	38	100.0	564	2 I61771	keratin 6f, type I
18	38	100.0	581	1 KRWS2	keratin, type II c
19	38	100.0	590	2 A29904	keratin 5, type II
20	38	100.0	629	2 A29666	keratin, 65K type
21	38	100.0	629	2 S42629	keratin K3 - rabbi
22	38	100.0	638	2 I53169	cyto keratin 2 - hu
23	38	100.0	643	1 KRHU2	keratin 1, type II
24	38	100.0	645	2 A44861	keratin, 67K type
25	36	94.7	681	2 T14903	NADPH-ferrihemopro
26	35	92.1	94	2 C59094	hypothetical prote
27	35	92.1	224	2 S00959	hypothetical prote
28	35	92.1	469	2 S05602	keratin K7, type I
29	35	92.1	489	2 B24177	keratin, 55K type

ALIGNMENTS

RESULT 1

S43865

Cyto keratin 8, type II - potoroos (fragment)

C:Species: Potorous tridactylus (potoroo, long-nosed rat kangaroo)

C:Date: 20-Feb-1995 #sequence\_revision 06-Dec-1996 #text\_change 09-Jul-2004

C:Accession: S43865; S40172

C:Boettger, V.; Lane, E.B.

J. Mol. Biol. 235, 61-67, 1994

A:Title: A monoclonal antibody epitope on keratin 8 identified using a phage peptide lib

A:Reference number: S43865; MUID:94118290; PMID:7507169

A:Accession: S43865

A:Molecule type: mRNA

A:Residues: 1-310 <BOE>

A:Cross-references: UNIPROT:Q28810; UNIPARC:UPI0000086BF0; EMBL:X70987; NID:9437902; PIDN:

A:Experimental source: kidney epithelium cells

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 100.0%; Score 38; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 4.1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7

DB 180 YEDEINK 186

RESULT 2

I61769

keratin 6d, type II - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C:Accession: I61769

R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.

J. Biol. Chem. 270, 18581-18592, 1995

A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A:Reference number: A57398; MUID:95355491; PMID:7543104

A:Accession: I61769

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-384 <RES>

A:Cross-references: UNIPROT:P48667; UNIPARC:UPI000012DB17; GB:L42602; NID:9914823; GB:L42

28; GB:L42608; NID:9914829; GB:L42609; NID:9914830; GB:L42610; NID:91488252; PIDN:AAB606

C:Genetics:

A:Gene: KRT6D

A:Introns: 72/2; 92/3; 124/3; 179/3; 221/3; 295/2; 307/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 100.0%; Score 38; DB 2; Length 384;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 YEDEINK 7
Db      73 YEDEINK 79

RESULT 3
A60093
cytokeratin, type II, early ectodermal - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 30-Sep-1993
C:Accession: A60093
R:Charlebois, T.S.; Spencer, D.H.; Tarkington, S.K.; Henry, J.J.; Grainger, R.M.
Development 108, 33-45, 1990
A:Title: Isolation of a chick cytokeratin cDNA clone indicative of regional specialization
A:Reference number: A60093; MUID:90276238; PMID:1693557
A:Accession: A60093
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-390 <CHA>
A:Cross-references: UNIPARC:UPI00001774C6
C:Superfamily: cytoskeletal keratin
C:Keywords: intermediate filament

Query Match      100.0%; Score 38; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 5.3; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy      1 YEDEINK 7
Db     100 YEDEINK 106

RESULT 4
KRXL2B
keratin, 64K type II cytoskeletal (clone pUF164) - African clawed frog (fragment)
N:Alternate names: 64-kDa type II keratin
C:Species: Xenopus laevis (African clawed frog)
C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 09-Jul-2004
C:Accession: A02952
R:Hoffmann, W.; Franz, J.K.; Franke, W.W.
J. Mol. Biol. 184, 713-724, 1985
A:Title: Amino acid sequence microheterogeneities of basic (type II) cytokeratins of Xenopus laevis
A:Reference number: A92914; MUID:86011576; PMID:2413219
A:Accession: A02952
A:Molecule type: mRNA
A:Residues: 1-419 <HOF>
A:Cross-references: UNIPROT:P04265; UNIPARC:UPI000012DB07; GB:X02895; GB:M13954; NID:G64
A:Experimental source: clone pUF164
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-275/Domain: rod (fragment) <ROD>
F:1-9/Region: linker 1
F:10-110/Region: coil 1B
F:111-127/Region: linker 12
F:128-146/Region: coil 2A
F:147-154/Region: linker 2
F:155-275/Region: coil 2B
F:213/Region: scutter
F:276-419/Domain: tail <END>
F:276-399/Region: H2 and V2 subdomains
F:400-419/Region: E2 subdomain

Query Match      100.0%; Score 38; DB 1; Length 419;
Best Local Similarity 100.0%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy      1 YEDEINK 7
Db     52 YEDEINK 58

RESULT 5

```

```

KRXL2A
keratin, 64K type II cytoskeletal (clone pUF23) - African clawed frog (fragment)
N:Alternate names: 64-kDa type II keratin
C:Species: Xenopus laevis (African clawed frog)
C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 09-Jul-2004
C:Accession: A02953
R:Hoffmann, W.; Franz, J.K.; Franke, W.W.
J. Mol. Biol. 184, 713-724, 1985
A:Title: Amino acid sequence microheterogeneities of basic (type II) cytokeratins of Xenopus laevis
A:Reference number: A92914; MUID:86011576; PMID:2413219
A:Accession: A02953
A:Molecule type: mRNA
A:Residues: 1-425 <HOF>
A:Cross-references: UNIPROT:P04266; UNIPARC:UPI000012DB06; GB:X02894; GB:M13956; NID:G64
A:Experimental source: clone pUF23
C:Comment: The cytokeratin IF proteins appear to be obligate heteropolymers, composed of types of proteins is low and restricted to specific isolated regions within the rod domain.
C:Comment: Three subdomains, each characterized by a distinctive amino acid composition, d length among the intermediate filament proteins, the arrangement of its subdomains pre-
ally distant species, suggests functional significance.
C:Comment: There is only one group of closely related type II amphibian cytokeratins. This
ate of differentiation of the amphibian epidermis, which lacks a true stratum corneum.
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-296/Domain: rod (fragment) <ROD>
F:1-16/Region: coil 1A (fragment)
F:11-30/Region: linker 1
F:31-131/Region: coil 1B
F:132-148/Region: linker 12
F:149-167/Region: coil 2A
F:168-175/Region: linker 2
F:176-296/Region: coil 2B
F:234/Region: scutter
F:297-425/Domain: tail <END>
F:297-405/Region: H2 and V2 subdomains
F:406-425/Region: E2 subdomain

Query Match      100.0%; Score 38; DB 1; Length 425;
Best Local Similarity 100.0%; Pred. No. 5.8; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy      1 YEDEINK 7
Db     73 YEDEINK 79

RESULT 6
A43782
keratin, type II - eastern newt (fragment)
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C:Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 09-Jul-2004
C:Accession: A43782
R:Ferretti, P.; Brockes, J.P.; Brown, R.
Development 111, 497-507, 1991
A:Title: A new type II keratin restricted to normal and regenerating limbs and tails is
A:Reference number: A43782; MUID:91372147; PMID:1716554
A:Accession: A43782
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-461 <FER>
A:Cross-references: UNIPROT:O42435; UNIPARC:UPI00001774C4; GB:X57671
C:Superfamily: cytoskeletal keratin

Query Match      100.0%; Score 38; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 6.3; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy      1 YEDEINK 7
Db     134 YEDEINK 140

RESULT 7

```



A34720  
keratin 8, type II cytoskeletal - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Jul-1990 #sequence revision 06-Jul-1990 #text change 09-Jul-2004  
C:Accession: A34720; S06888; S12479; I37982; I37983; JS0487; A31942  
R:Yamamoto, R.; Kao, L.C.; McKnight, C.E.; Strauss III, J.F.  
Mol. Endocrinol. 4, 370-374, 1990  
A:Title: Cloning and sequence of cDNA for human placental cytokeratin 8. Regulation of t  
A:Reference number: A34720; MUID:90258929; PMID:1692965  
A:Accession: A34720  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-483 <YAM>  
A:Cross-references: UNIPROT:P05787; UNIPARC:UPI00001774C3; GB:M34225  
R:Leube, R.E.; Bosch, F.X.; Romano, V.; Zimbelmann, R.; Hoesler, H.; Franke, W.W.  
Differentiation 33, 69-85, 1986  
A:Title: Cytokeratin expression in simple epithelia.  
A:Reference number: S06888; MUID:87134779; PMID:2434381  
A:Accession: S06888  
A:Molecule type: mRNA  
A:Residues: 205-483 <LEU>  
A:Cross-references: UNIPARC:UPI000016ABAI; EMBL:X12882  
R:Franke, W.W.  
submitted to the EMBL Data Library, September 1988  
A:Reference number: S12479  
A:Accession: S12479  
A:Molecule type: mRNA  
A:Residues: 205-309, '1', 311-483 <FRA>  
A:Cross-references: UNIPARC:UPI000016A769; EMBL:X12882; NID:G30312; PIDN:CAA31376.1; PID  
R:Wasseem, A.; Alexander, C.M.; Steel, J.B.; Lane, E.B.  
New Biol. 2, 464-478, 1990  
A:Title: Embryonic simple epithelial keratins 8 and 18: chromosomal location emphasizes  
A:Reference number: I37982; MUID:91145351; PMID:1705144  
A:Accession: I37982  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-76, 'V', 78-428, 'G', 430-431, 'S', 433-483 <RES>  
A:Cross-references: UNIPARC:UPI0000161B3E; EMBL:X74929; NID:G400415; PIDN:CAAS2882.1; PI  
A:Accession: I37983  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 151-201, 'VD', 202-271 <RE2>  
A:Cross-references: UNIPARC:UPI000016ABCF; EMBL:X74981; NID:G400417; PIDN:CAAS2916.1; PI  
R:Krauss, S.; Franke, W.W.  
Gene 86, 241-249, 1990  
A:Title: Organization and sequence of the human gene encoding cytokeratin 8.  
A:Reference number: JS0487; MUID:90215304; PMID:1691124  
A:Accession: JS0487  
A:Molecule type: DNA  
A:Residues: 1-76, 'V', 78-428, 'G', 'SQA', 434-483 <KRA>  
A:Cross-references: UNIPARC:UPI000016A82B; GB:M34482; NID:G181572; PIDN:AAA35763.1; PID:  
R:Kulesh, D.A.; Cecena, G.; Darmon, Y.M.; Vasseur, M.; Oshima, R.G.  
Mol. Cell. Biol. 9, 1553-1565, 1989  
A:Title: Posttranslational regulation of keratins: degradation of mouse and human kerati  
A:Reference number: A31942; MUID:89261783; PMID:2471065  
A:Accession: A31942  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-76, 'V', 78-231 <KUL>  
A:Cross-references: UNIPARC:UPI000016A354; GB:M26512; NID:G177796; PIDN:AAA51542.1; PID:  
A:Comment: This protein is one of the type II cytokeratins.  
C:Genetics:  
A:Gene: GDB:KRT8  
A:Cross-references: GDB:118830; OMIM:148060  
A:Map position: 12p13.2-12q24.1  
A:Introns: 108/3; 178/2; 198/3; 230/3; 327/3; 401/2; 421/1  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; intermediate filament

Query Match 100.0%; Score 38; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
Db 179 YEDEINK 185

RESULT 8  
J0407  
keratin 8, type II cytoskeletal, embryonic - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C:Accession: J0407; A22078  
R:Ouillet, T.; Levac, P.; Royal, A.  
Gene 70, 75-84, 1988  
A:Title: Complete sequence of the mouse type-II keratin EndoA: its amino-terminal region  
A:Reference number: J0407; MUID:89196919; PMID:2467842  
A:Accession: J0407  
A:Molecule type: mRNA  
A:Residues: 1-487 <OUE>  
A:Cross-references: UNIPROT:Q61518; UNIPARC:UPI00000E6EE0; GB:M22831; NID:G511653; PIDN:  
A:Experimental source: carcinoma cell line F9-21  
R:Vasseur, M.; Duprey, P.; Brulet, P.; Jacob, F.  
Proc. Natl. Acad. Sci. U.S.A. 82, 1155-1159, 1985  
A:Title: One gene and one pseudogene for the cytokeratin endo A.  
A:Reference number: A22078; MUID:85140247; PMID:2579387  
A:Accession: A22078  
A:Molecule type: DNA  
A:Residues: 1-13 <VAS>  
A:Cross-references: UNIPARC:UPI000016CD31; GB:K02783; NID:G193019; PIDN:AAA37549.1; PID:  
C:Genetics:  
A:Gene: EndoA  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; intermediate filament  
F:1-94/Domain: head <HED>  
F:95-406/Domain: rod <ROD>  
F:95-128/Region: coil 1A  
F:129-146/Region: linker 1  
F:147-235/Region: coil 1B  
F:236-260/Region: linker 12  
F:261-406/Region: coil 2  
F:407-487/Domain: tail <TAI>

Query Match 100.0%; Score 38; DB 2; Length 487;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
Db 182 YEDEINK 188

RESULT 9  
J0658  
cytokeratin EndoA - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C:Accession: JS0658  
R:Tamai, Y.; Takemoto, Y.; Matsumoto, M.; Morita, T.; Matsushiro, A.; Nozaki, M.  
Gene 104, 169-176, 1991  
A:Title: Sequence of the EndoA gene encoding mouse cytokeratin and its methylation state  
A:Reference number: JS0658; MUID:92009210; PMID:1717348  
A:Accession: JS0658  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-490 <TAM>  
A:Cross-references: UNIPROT:Q61463; UNIPARC:UPI00000220F8; GB:D90360; NID:G220391; PIDN:  
A:Comment: This type-II cytokeratin forms intermediate filaments together with type-I cy  
C:Genetics:  
A:Gene: EndoA  
A:Introns: 114/3; 184/2; 204/3; 236/3; 291/3; 333/3; 407/2; 427/1  
C:Superfamily: cytoskeletal keratin

Query Match 100.0%; Score 38; DB 2; Length 490;  
Best Local Similarity 100.0%; Pred. No. 6.7;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7  
 |||||  
 Db 185 YEDEINK 191

## RESULT 10

S08381  
 keratin, 58K type II, cytoskeletal (clone pXenCK55(5/6)) - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 09-Jul-2004  
 C:Accession: S08381  
 R:Fouquet, B.; Herrmann, H.; Franz, J.K.; Franke, W.W.  
 Development 104, 533-548, 1998  
 A:Title: Expression of intermediate filament proteins during development of Xenopus laevis  
 A:Reference number: S08381; MUID:90032362; PMID:2478354  
 A:Accession: S08381  
 A:Molecule type: mRNA  
 A:Residues: 1-513 <FOU>  
 A:Cross-references: UNIPROT:P16978; UNIPARC:UPI000017151D; EMBL:X14427; NID:G64625; PIDN:  
 C:Superfamily: cytoskeletal keratin

Query Match 100.0%; Score 38; DB 2; Length 513;  
 Best Local Similarity 100.0%; Pred. No. 7.1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7  
 |||||  
 Db 217 YEDEINK 223

## RESULT 11

JS0291  
 intermediate filament protein ON3 - goldfish  
 C:Species: Carassius auratus (goldfish)  
 C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 09-Jul-2004  
 C:Accession: JS0291  
 R:Giordano, S.; Glasgow, E.; Tesser, P.; Schechter, N.  
 Neuron 2, 1507-1516, 1989  
 A:Title: A type II keratin is expressed in glial cells of the goldfish visual pathway.  
 A:Reference number: JS0291; MUID:90180470; PMID:2483326  
 A:Accession: JS0291  
 A:Molecule type: mRNA  
 A:Residues: 1-520 <GIO>  
 A:Cross-references: UNIPROT:P18520; UNIPARC:UPI000012D74A  
 A:Experimental source: glial cell  
 C:Comment: This protein is one of the nonneuronal predominant intermediate filament proteins  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; optic nerve; visual pathway  
 F:1-108/Domain: head <HEA>  
 F:109-144/Region: coil 1A  
 F:145-157/Region: linker  
 F:158-253/Region: coil 1B  
 F:254-273/Region: linker  
 F:274-420/Region: coil 2  
 F:421-520/Domain: tail <TAI>

Query Match 100.0%; Score 38; DB 2; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 7.2; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7  
 |||||  
 Db 197 YEDEINK 203

## RESULT 12

I59009  
 epidermal keratin subunit II - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1995 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
 C:Accession: I59009; A61205

R:Steinert, P.M.; Parry, D.A.D.; Racoosin, E.L.; Idler, W.W.; Steven, A.C.; Trus, B.L.; Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713, 1984  
 A:Title: The complete cDNA and deduced amino acid sequence of a type II mouse epidermal keratin  
 A:Reference number: I59009; MUID:85014838; PMID:6207530  
 A:Accession: I59009  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-553 <RES>  
 A:Cross-references: UNIPROT:P50446; UNIPARC:UPI000016CE87; GB:K02108; NID:G198634; PIDN:I  
 R:Finch, J.; Andrews, K.; Krieg, P.; Fuerstenberger, G.; Slaga, T.; Ootsuyama, A.; Tanoh  
 Carcinogenesis 12, 1519-1522, 1991  
 A:Title: Identification of a cloned sequence activated during multi-stage carcinogenesis  
 A:Reference number: A61205; MUID:91316763; PMID:1713533  
 A:Accession: A61205  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 528-553 <FIN>  
 A:Cross-references: UNIPARC:UPI00001774C8  
 C:Genetics:  
 A:Gene: KER2  
 C:Superfamily: cytoskeletal keratin

Query Match 100.0%; Score 38; DB 2; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 7.7; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7  
 |||||  
 Db 242 YEDEINK 248

## RESULT 13

KRHUEA  
 keratin 6a, type II - human  
 N:Alternate names: 56-kDa type II keratin; keratin cytoskeletal  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Nov-1984 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: A57398; A02944  
 R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
 J. Biol. Chem. 270, 18581-18592, 1995  
 A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
 A:Reference number: A57398; MUID:95355491; PMID:7543104  
 A:Accession: A57398  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-564 <TAK>  
 A:Cross-references: UNIPROT:P02538; UNIPARC:UPI000013CD4C; GB:I42575; NID:G908769; GB:I42  
 74; GB:I42581; NID:G908775; GB:I42582; NID:G908776; GB:I42583; NID:G908777; PIDN:AAC4176;  
 R:Hanukoglu, I.; Fuchs, E.  
 Cell 33, 915-924, 1983  
 A:Title: The cDNA sequence of a type II cytoskeletal keratin reveals constant and variab  
 A:Reference number: A02944; MUID:83259278; PMID:6191871  
 A:Accession: A02944  
 A:Molecule type: mRNA  
 A:Residues: 208-394, 'S', 396-564 <HAN>  
 A:Cross-references: UNIPARC:UPI000016AB87; GB:J00269; NID:G34068; PIDN:CAA24760.1; PID:G  
 C:Comment: The cytoskeletal and microfibrillar keratins are classified into two types, ty  
 atin IF protein subunit appears to be a heterotetramer of two type I and two type II prot  
 C:Genetics:  
 A:Gene: GDB:KRT6A  
 A:Cross-references: GDB:I28111; OMIM:148041  
 A:Map position: 12q12-12q21  
 A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; intermediate filament  
 F:2-163/Domain: head <HED>  
 F:127-163/Region: E1 and V1 subdomains  
 F:162-476/Domain: rod <ROD>  
 F:199-210/Region: linker 1  
 F:211-311/Region: coil 1B  
 F:312-328/Region: linker 12  
 F:329-347/Region: coil 2A

F:348-355/Region: linker 2  
F:356-476/Region: coil 2B  
F:414/Region: stutter  
F:477-564/Domain: tail <END>  
F:477-496/Region: H2 subdomain  
F:497-564/Region: V2 and E2 subdomains

Query Match 100.0%; Score 38; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
|||  
DB 253 YEDEINK 259

## RESULT 14

KRUEB

keratin 6b, type II - human  
N:Alternate names: 56K type II keratin; keratin, cytoskeletal  
C:Species: Homo sapiens (man)  
C:Date: 04-Dec-1986 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: I61767; A02945  
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A:Reference number: A57398; MUID:95355491; PMID:7543104  
A:Accession: I61767

A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-564 <RES>

A:Cross-references: UNIPROT:P04259; UNIPARC:UPI0000167B38; GB:L42584; NID:9908780; GB:L4  
85; GB:L42590; NID:9908786; GB:L42591; NID:9908787; GB:L42592; NID:9908788; PIDN:AA4176  
R:Tyner, A.L.; Eichman, M.J.; Fuchs, E.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4683-4687, 1985

A:Title: The sequence of a type II keratin gene expressed in human skin: conservation of  
A:Reference number: A02945; MUID:85270392; PMID:2410904  
A:Accession: A02945

A:Molecule type: DNA  
A:Residues: 2-88, 'AG', '91-115, 'PA', '118', '122-158, 'IG', '161-254, 'V', '256-564 <TYN>  
A:Cross-references: UNIPARC:UPI0000173D58; GB:M11229; GB:L00205; NID:g186714; PIDN:AAA59

A:Note: initiator Met not shown  
C:Comment: There are two types of cytoskeletal and microfibrillar keratin, I (acidic) and  
bi-gate heteropolymers, composed of complexes formed by the aggregation of at least one  
C:Genetics:

A:Gene: GDB:KRT6B

A:Cross-references: GDB:128113; OMIM:148042

A:Map position: 12pter-12qter

A:Introns: 180/3; 252/3; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

F:2-163/Domain: head <HED>

F:2-127/Region: E1 and V1 subdomains

F:128-163/Region: H1 subdomain

F:164-476/Domain: rod <ROD>

F:164-198/Region: coil 1A

F:199-210/Region: linker 1

F:211-311/Region: coil 1B

F:312-328/Region: linker 12

F:329-347/Region: coil 2A

F:348-355/Region: linker 2

F:356-476/Region: coil 2B

F:414/Region: stutter

F:477-564/Domain: tail <END>

F:477-496/Region: H2 subdomain

F:497-564/Region: V2 and E2 subdomains

Query Match 100.0%; Score 38; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
|||  
DB 253 YEDEINK 259

Db 253 YEDEINK 259

## RESULT 15

I61770

keratin 6e, type II - human  
C:Species: Homo sapiens (man)

C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C:Accession: I61770

R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.

J. Biol. Chem. 270, 18581-18592, 1995

A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A:Reference number: A57398; MUID:95355491; PMID:7543104

A:Accession: I61770

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-564 <RES>

A:Cross-references: UNIPROT:P48668; UNIPARC:UPI0000167E31; GB:L42611; NID:9908802; PIDN:7

C:Genetics:

A:Gene: KRT6E

A:Note: this may not be a distinct gene

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 100.0%; Score 38; DB 2; Length 564;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
|||  
DB 253 YEDEINK 259

Search completed: July 5, 2006, 19:20:30  
Job time: 8.49515 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 71.6311 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-1  
Perfect score: 38  
Sequence: 1 YEDEINK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	38	100.0	176	2	Q4JI23 ICTPU	Q4ji23 ictalurus p
2	38	100.0	222	2	Q4L0E7 MOUSE	Q4l0e7 mus musculus
3	38	100.0	224	2	Q678L1 MOUSE	Q678l1 mus musculus
4	38	100.0	310	2	Q28810 POTTR	Q28810 pottorus tr
5	38	100.0	324	2	Q80VP7 MOUSE	Q80vp7 mus musculus
6	38	100.0	346	2	Q4QY72 SPAAU	Q4qy72 sparus aura
7	38	100.0	365	2	Q6910 HUMAN	Q6910 homo sapien
8	38	100.0	381	2	Q99NH7 MOUSE	Q99nh7 mus musculus
9	38	100.0	384	1	K2C6D HUMAN	P48667 homo sapien
10	38	100.0	417	2	Q4SRP0 TETNG	Q4srp0 tetraodon n
11	38	100.0	419	1	K2C2 XENLA	P04265 xenopus lae
12	38	100.0	425	1	K2C1 XENLA	P04266 xenopus lae
13	38	100.0	461	2	Q42435 NOTVI	Q42435 notophthalm
14	38	100.0	482	1	K2C8 HUMAN	P05787 homo sapien
15	38	100.0	482	1	K2C8 RAT	Q10758 rattus norv
16	38	100.0	483	2	Q53GJ0 HUMAN	Q53gj0 homo sapien
17	38	100.0	483	2	Q6DHW5 HUMAN	Q6dhw5 homo sapien
18	38	100.0	483	2	Q6GMY0 HUMAN	Q6gmy0 homo sapien
19	38	100.0	483	2	Q5RB34 PONPY	Q5rb34 pongo pygma
20	38	100.0	483	2	Q80WH8 SMURI	Q80wh8 rattus sp.
21	38	100.0	486	2	Q9YH07 9CHON	Q9yh07 scylliorhinu
22	38	100.0	487	2	Q61518 MOUSE	Q61518 mus musculus
23	38	100.0	487	2	Q7SYF6 ACIBE	Q7syf6 acipenser b
24	38	100.0	490	2	Q3KQK5 MOUSE	Q3kqk5 mus musculus
25	38	100.0	490	2	Q3TG11 MOUSE	Q3tg11 mus musculus
26	38	100.0	490	2	Q3TJEL MOUSE	Q3tjel mus musculus
27	38	100.0	490	2	Q3TKY7 MOUSE	Q3tky7 m blastocys
28	38	100.0	490	2	Q61463 MOUSE	Q61463 m cytokerat
29	38	100.0	492	1	K2CO CHICK	Q93532 gallus gall
30	38	100.0	496	2	Q42434 NOTVI	Q42434 notophthalm
31	38	100.0	497	2	Q5K2N3_PROAT	Q5k2n3 protocerperus

32	38	100.0	497	2	Q6NY60 BRARE	Q6ny60 brachydanio
33	38	100.0	498	2	Q9PV92 BRARE	Q9pv92 brachydanio
34	38	100.0	499	2	Q9H552 HUMAN	Q9h552 homo sapien
35	38	100.0	499	2	Q7ZT78 BRARE	Q7zt78 brachydanio
36	38	100.0	505	2	Q6P4C7 HUMAN	Q6p4c7 homo sapien
37	38	100.0	508	2	Q6NVR6_XENTR	Q6nvr6 xenopus tro
38	38	100.0	511	2	Q90ZF7_RANCA	Q90zf7 rana catesb
39	38	100.0	512	1	K2CS_XENLA	P16878 xenopus lae
40	38	100.0	515	2	Q8JFG4 ONCMY	Q8jfg4 oncorhynchu
41	38	100.0	519	1	K2C1B RAT	Q6i01 rattus norv
42	38	100.0	520	1	ION3 CARAU	P18520 carassius a
43	38	100.0	520	2	Q6NWF6 BRARE	Q6nwf6 brachydanio
44	38	100.0	521	2	Q5K2N4_PROAT	Q5k2n4 protocerperus
45	38	100.0	521	2	Q5K2N5_PROAT	Q5k2n5 protocerperus

ALIGNMENTS

RESULT 1  
Q4JI23 ICTPU PRELIMINARY; PRT; 176 AA.  
AC Q4JI23;  
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 02-AUG-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Keratin 8 (Fragment).  
OS Ictalurus punctatus (Channel catfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
OC Ictaluridae; Ictalurus.  
OX NCBI\_TaxID=7998;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Yeh H.-Y., Klesius P.H.;  
RT "Differential gene expression of Ictalurus punctatus after  
RT Edwardsiella ictaluri infection."  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC EMBL; DO086174; AAY86957.1; --; mRNA.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002957; Keratin I.  
DR InterPro; IPR003054; Keratin II.  
DR PANTHER; PTHR18893:SP5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PRINTS; PR01276; TYPE2KERATIN.  
KW Keratin.  
FT NON\_TER 1  
FT NON\_TER 176 176  
SQ SEQUENCE 176 AA; 20651 MW; 8EE3A88097529C33 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
|||  
DB 70 YEDEINK 76

RESULT 2  
Q4L0E7 MOUSE PRELIMINARY; PRT; 222 AA.  
AC Q4L0E7;  
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 02-AUG-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Type II cytokeratin Kb39 (Fragment).

```

GN Name=4732484G22Rik;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NMRI; TISSUE=Tail skin;
RA Zinek A., Heese M.;
RT "Cornea-specific cytochrome B3 seems to be absent in the rodent
lineage.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY651826; AAV85505.1; -; mRNA.
DR MGI; MGI:3588209; 4732484G22Rik.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; Filament; 1.
DR PANTHER; PTHR1893.SF5; Keratin_II; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
FT NON TER 1
FT NON TER 222
FT NON TER 222
SQ SEQUENCE 222 AA; 25801 MW; 483B4F589032C1D1 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 222;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 YEDEINK 7
Db 70 YEDEINK 76

RESULT 3
Q678L1 MOUSE
ID Q678L1 MOUSE PRELIMINARY; PRT; 224 AA.
AC Q678L1
DT 11-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Keratin 1b (Fragment).
DN Name=4732484G22Rik;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NMRI; TISSUE=Tail skin;
RA Zinek A., Heese M., Magin T., Weber K.;
RT "Homologies between the human, mouse and rat keratin-clusters.";
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY379112; AAR33930.1; -; mRNA.
DR Ensembl; ENSMUSG0000067594; Mus musculus.
DR MGI; MGI:3588209; 4732484G22Rik.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.

Query Match 100.0%; Score 38; DB 2; Length 222;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 YEDEINK 7
Db 70 YEDEINK 76

RESULT 3
Q678L1 MOUSE
ID Q678L1 MOUSE PRELIMINARY; PRT; 224 AA.
AC Q678L1
DT 11-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Keratin 1b (Fragment).
DN Name=4732484G22Rik;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NMRI; TISSUE=Tail skin;
RA Zinek A., Heese M., Magin T., Weber K.;
RT "Homologies between the human, mouse and rat keratin-clusters.";
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY379112; AAR33930.1; -; mRNA.
DR Ensembl; ENSMUSG0000067594; Mus musculus.
DR MGI; MGI:3588209; 4732484G22Rik.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.

```

```

DR PANTHER; PTHR1893.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
FT NON TER 1
FT NON TER 224
FT NON TER 224
SQ SEQUENCE 224 AA; 25995 MW; DD6F783B4F589032 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 224;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 YEDEINK 7
Db 70 YEDEINK 76

RESULT 4
Q28810 POTTR
ID Q28810 POTTR PRELIMINARY; PRT; 310 AA.
AC Q28810
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Keratin 8 (Fragment).
OS Potorous tridactylus (Potoroo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Potoroidae; Potorous.
OX NCBI_TaxID=9310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94118290; Pubmed=7507169;
RA Boettger V., Lane E.B.;
RT "A monoclonal antibody Epitope on Keratin 8 identified using a phage
peptide library.";
RL J. Mol. Biol. 235:61-67 (1994).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; X70987; CAA50316.1; -; mRNA.
DR PIR; S43865; S43865.
DR HSSP; P08670; 1GK7.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
FT NON TER 1
FT NON TER 310
FT NON TER 310
SQ SEQUENCE 310 AA; 35170 MW; 31D37E0BC63BFA85 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 310;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 YEDEINK 7
Db 180 YEDEINK 186

RESULT 5
Q80VP7 MOUSE
ID Q80VP7 MOUSE PRELIMINARY; PRT; 324 AA.
AC Q80VP7
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.

```

DT	07-FEB-2006, entry version 20.
DE	Hypothetical protein MGC54654.
DN	Name=MGC54654;
GC	Mus musculus (Mouse).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC	Muroidea; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RP	[1]
RN	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Eye;
RC	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA	Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA	Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.
RA	Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences.";
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]	
RN	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Eye;
RG	NIH MGC Project;
RG	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	EMBL; BC046626; AH46626.1; -; mRNA.
DR	HSSP; P08670; IGK7.
DR	Ensembl; ENSMUSG00000046162; Mus musculus.
DR	GO; GO:0005198; F:structural molecule activity; IEA.
DR	InterPro; IPR001664; IF.
DR	InterPro; IPR003054; Keratin II.
DR	PANTHER; PTHR18993:SF5; Keratin_II; 1.
DR	Pfam; PF00038; Filament; 1.
DR	PRINTS; PR01276; TYPE2KERATIN.
KW	Hypothetical protein.
SW	SEQUENCE 324 AA; 35165 MW; 0E4D7E25144435DD CRC64;
Query Match	100.0%; Score 38; DB 2; Length 324;
Best Local Similarity	100.0%; Pred. No. 50;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	1 YEDEINK 7
DB	203 YEDEINK 209
RESULT 6	
ID	Q4QY72_SPAAU PRELIMINARY; PRT; 346 AA.
AC	Q4QY72;
DT	19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT	19-JUL-2005, sequence version 1.
DT	07-FEB-2006, entry version 6.
DE	Type II keratin E3-like protein.
OS	Sparus aurata (Gilthead sea bream).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RG NIH MGC Project;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: BC011373; AAH11373.2; -; mRNA.
DR Ensembl: ENSG00000170421; Homo sapiens.
DR GO: GO:0005882; C:intermediate filament; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR InterPro: IPR001664; IF.
DR InterPro: IPR002957; Keratin_I.
DR PANTHER: PTHR1893:SP5; Keratin_II; 1.
DR Pfam: PF00038; Filament; 1.
DR PRINTS: PRO1248; TYPE1KERATIN.
DR PROSITE: PRO1276; TYPE2KERATIN.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament.
FT NON TER 1
SQ SEQUENCE 365 AA; 41108 MW; 57CD83091D0635E9 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 61 YEDEINK 67

RESULT 8
Q99MH7 MOUSE PRELIMINARY; PRT; 381 AA.
ID Q99MH7;
AC Q99MH7;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE Cyokeratin KRT2-6HF (Fragment).
GN Name=Krtcpl; Synonyms=Krt2-6hf;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C3H/HeN;
RX MEDLINE=22286385; PubMed=12399393;
RA Polier C., Yoshiki A., Fujiwara K., Guenet J.-L., Kusakabe M.;
RT "Haque (Hag): a new mouse hair mutation with an unstable semidominant
RT allele.";
RL Genetics 162:831-840 (2002).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AF343088; AAK17206.1; -; mRNA.
DR HSSP: P08670; IGH4.
DR Ensembl: ENSMUSG00000022986; Mus musculus.
DR MGI: MGI:1923500; Krtcpl.
DR GO: GO:0005882; C:intermediate filament; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR InterPro: IPR001664; IF.
DR InterPro: IPR002957; Keratin_I.
DR InterPro: IPR003054; Keratin_II.

```

```

DR PANTHER: PTHR1893:SP5; Keratin_II; 1.
DR Pfam: PF00038; Filament; 1.
DR PRINTS: PRO1248; TYPE1KERATIN.
DR PROSITE: PRO1276; TYPE2KERATIN.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Keratin.
FT NON TER 1
SQ SEQUENCE 381 AA; 42383 MW; 4143A882F7F6E9A1 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 70 YEDEINK 76

RESULT 9
K2C6D HUMAN STANDARD; PRT; 384 AA.
ID K2C6D_HUMAN
AC K2C6D;
DT 01-FEB-1996, integrated into UniProtKB/Swiss-Prot.
DT 01-FEB-1996, sequence version 1.
DT 07-FEB-2006, entry version 44.
DE Keratin, type II Cytoskeletal 6D (Cytokeratin-6D) (CK 6D) (K6d
DE keratin) (Fragment).
GN Name=KRT6d;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC TISSUE=Skin;
RX MEDLINE=95355491; PubMed=7543104; DOI=10.1074/jbc.270.31.18581;
RA Takahashi K., Paladini R.D., Coulombe P.A.;
RT "Cloning and characterization of multiple human genes and cDNAs
RT encoding highly related type II keratin 6 isoforms.";
RL J. Biol. Chem. 270:18581-18592 (1995).
CC -!- FUNCTION: There are two types of cytoskeletal and microfibrillar
CC keratin: I (acidic; 40-55 kDa) [K9 to K20] and II (neutral to
CC basic; 56-70 kDa) [K1 to K8]. Both a basic and an acidic keratin
CC are required for filament assembly.
CC -!- SUBUNIT: Heterodimer of a type I and a type II keratin. KRT6
CC isomers associate with KRT16 and/or KRT17.
CC -!- TISSUE SPECIFICITY: Constitutively expressed in distinct types of
CC epithelia such as those in oral mucosa, esophagus, papillae of
CC tongue and hair follicle outer root sheath.
CC -!- MISCELLANEOUS: There are at least six isoforms of human type II
CC keratin 6 (K6).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: L42610; AAB60696.1; -; Genomic_DNA.
DR EMBL: L42603; AAB60696.1; JOINED; Genomic_DNA.
DR EMBL: L42604; AAB60696.1; JOINED; Genomic_DNA.
DR EMBL: L42605; AAB60696.1; JOINED; Genomic_DNA.
DR EMBL: L42606; AAB60696.1; JOINED; Genomic_DNA.
DR EMBL: L42607; AAB60696.1; JOINED; Genomic_DNA.
DR EMBL: L42608; AAB60696.1; JOINED; Genomic_DNA.
DR EMBL: L42609; AAB60696.1; JOINED; Genomic_DNA.
DR PIR: I61769; I61769.
DR HSSP: P08670; IGH4.
DR Ensembl: ENSG00000170465; Homo sapiens.
DR HGNC: HGNC:20405; KRT6D.
DR GO: GO:0005200; F:structural constituent of cytoskeleton; NAS.
DR GO: GO:0030154; P:cell differentiation; NAS.
DR GO: GO:0008284; P:positive regulation of cell proliferation; NAS.
DR InterPro: IPR001664; IF.

```



```
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament; Keratin.
FT CHAIN <1 384 Keratin, type II cytoskeletal 6D.
FT REGION <1 292 Rod.
FT REGION <1 18 Coil 1A.
FT REGION 19 37 Linker 1.
FT REGION 38 129 Coil 1B.
FT REGION 130 153 Linker 12.
FT REGION 154 292 Coil 2.
FT REGION 293 384 Tail.
FT SITE 234 234 Stutter.
FT NON_TER 1 1
SQ SEQUENCE 384 AA; 42468 MW; D69DCF0808C94E91 CRC64;

Query Match 100.0%; Score 38; DB 1; Length 384;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 73 YEDEINK 79

RESULT 10
O4SRP0_TETNG PRELIMINARY; PRT; 417 AA.
ID Q4SRP0_TETNG
AC Q4SRP0_TETNG
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-FEB-2006, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Chromosome undetermined SCAR14505, whole genome shotgun sequence.
GN ORFNames=GSTENG00013817001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Winkler P., Lander E.S., Weissbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; CAAE01014505; CAF96692.1; -; Genomic DNA.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament.
SQ SEQUENCE 417 AA; 46557 MW; B2F4D4E295625589 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 417;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 222 YEDEINK 228

RESULT 11
K2C2_XENLA
ID K2C2_XENLA STANDARD; PRT; 419 AA.
AC P04265;
DT 20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT 20-MAR-1987, sequence version 1.
DT 07-FEB-2006, entry version 39.
DE Keratin, type II cytoskeletal I (Clone PUF164) (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=86011576; PubMed=2413219;
RA Hoffmann W., Franz J.K., Franke W.W.;
RT "Amino acid sequence microheterogeneities of basic (type II)
RT cytokeratins of Xenopus laevis epidermis and evolutionary
RT conservativity of helical and non-helical domains.";
RL J. Mol. Biol. 184:713-724(1985).
CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and
CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
CC 55 and 56-70 kilodaltons, respectively).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; X02895; CAA26654.1; -; mRNA.
DR PIR; A02952; KRXL2B.
DR HSP; P08670; 1GK4.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament; Keratin.
FT CHAIN <1 419 Keratin, type II cytoskeletal I.
FT REGION <1 271 Rod.
FT REGION <1 16 Linker 1.
FT REGION 17 108 Coil 1B.
FT REGION 109 132 Linker 12.
FT REGION 133 271 Coil 2.
FT REGION 272 419 Tail.
FT SITE 215 215 Stutter.
FT NON_TER 1 1
SQ SEQUENCE 419 AA; 44595 MW; 15D79584FC31DF15 CRC64;
```

```

Query Match      100.0%; Score 38; DB 1; Length 419;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 52 YEDEINK 58

RESULT 12
K2C1 XENLA
ID K2C1 XENLA STANDARD; PRT; 425 AA.
AC P04266;
DT 20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT 20-MAR-1987, sequence version 1.
DE Keratin, type II cytoskeletal I (clone PUF23) (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=86011576; PubMed=2413219;
RA Hoffmann W., Franz J.K., Franke W.W.;
RT "Amino acid sequence microheterogeneities of basic (type II)
cytokeratins of Xenopus laevis epidermis and evolutionary
conservativity of helical and non-helical domains.";
RJ J. Mol. Biol. 184:713-724 (1985).
CC -!- SUBUNIT: Heterotetramer of two type I and two type II keratins.
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and
microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
55 and 56-70 kilodaltons, respectively).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; X02894; CAA26653.1; -; mRNA.
DR PIR; A02953; KRXL2A.
DR HSSP; P08670; 1GK4.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament; Keratin.
FT CHAIN <1 425 Keratin, type II cytoskeletal I.
FT /FTID=PRO_0000063718.
FT REGION <1 292 Rod.
FT REGION <1 16 Coil 1A.
FT REGION 17 37 Linker 1.
FT REGION 38 129 Coil 1B.
FT REGION 130 153 Linker 12.
FT REGION 154 292 Coil 2.
FT REGION 293 425 Tail.
FT SITE 236 236 Stutter.
FT NON TER 1
SQ SEQUENCE 425 AA; 45725 MW; D58D15DEC8C1C4E3 CRC64;

Query Match      100.0%; Score 38; DB 1; Length 425;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 73 YEDEINK 79

RESULT 13
O42435 NOTVI
ID O42435 NOTVI PRELIMINARY; PRT; 461 AA.
AC O42435;
DT 01-JAN-1998, integrated into UniProtKB/TREMBL.
DT 01-JAN-1998, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Cytokeratin type II (Fragment).
GN Name=K11;
OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
OC Notophthalmus.
OX NCBI_TaxID=8316;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Limb blastema;
RA MEDLINE=91372147; PubMed=1716554;
RX Ferretti P., Brookes J.P., Brown R.;
RT "A new type II keratin restricted to normal and regenerating limbs
and tails is responsive to retinoic acid.";
RL Development 111:497-507 (1991).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Limb blastema;
RA Ferretti P.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBSJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AJ001295; CAA04655.1; -; mRNA.
DR PIR; A43782; A43782.
DR HSSP; P08670; 1GK7.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
FT NON TER 1
SQ SEQUENCE 461 AA; 50638 MW; F3A018A7C10CA0FD CRC64;

Query Match      100.0%; Score 38; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 134 YEDEINK 140

RESULT 14
K2C8 HUMAN
ID K2C8 HUMAN STANDARD; PRT; 482 AA.
AC P05787; Q14099; Q14716; Q14717; Q96J60;
DT 01-NOV-1988, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 5.
DT 07-FEB-2006, entry version 72.
DE Keratin, type II cytoskeletal 8 (Cytokeratin-8) (CK-8) (Keraton-8)
(K8).
GN Name=KRT8; Synonyms=CYK8;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

RX MEDLINE=90258929; PubMed=1692965;  
RA Yamamoto R., Kao L.C., McKnight C.E., Strauss J.F. III;  
RT "Cloning and sequence of cDNA for human placental cytokeratin 8.  
RT Regulation of the mRNA in trophoblastic cells by cAMP.";  
RL Mol. Endocrinol. 4:370-374(1990).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=90215304; PubMed=1691124; DOI=10.1016/0378-1119(90)90285-Y;  
RA Krauss S., Franke W.W.;  
RT "Organization and sequence of the human gene encoding cytokeratin 8.";  
RL Gene 86:241-249(1990).  
RN [3]  
RP NUCLEOTIDE SEQUENCE  
RX MEDLINE=91145351; PubMed=1705144;  
RA Wassem A., Alexander C.M., Steel J.B., Lane E.B.;  
RT "Embryonic simple epithelial keratins 8 and 18: Chromosomal location  
RT emphasizes difference from other keratin pairs.";  
RL New Biol. 2:464-478(1990).  
RN [4]  
RP NUCLEOTIDE SEQUENCE, AND PHOSPHORYLATION SITES SER-23 AND SER-431.  
RX MEDLINE=97207326; PubMed=9054461; DOI=10.1074/jbc.272.11.7556;  
RA Ku N.-O., Omary M.B.;  
RT "Phosphorylation of human keratin 8 in vivo at conserved head domain  
RT serine 23 and at epidermal growth factor-stimulated tail domain serine  
RT 431.";  
RL J. Biol. Chem. 272:7556-7564(1997).  
RN [5]  
RP TISSUE=Colon;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalick D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [6]  
RP NUCLEOTIDE SEQUENCE OF 1-230.  
RX MEDLINE=89261783; PubMed=2471065;  
RA Kulesh D.A., Cecena G., Darmon Y.M., Vasseur M., Oshima R.G.;  
RT "Posttranslational regulation of keratins: degradation of mouse and  
RT human keratins 18 and 8.";  
RL Mol. Cell. Biol. 9:1553-1565(1989).  
RN [7]  
RP NUCLEOTIDE SEQUENCE OF 204-482.  
RX MEDLINE=87134779; PubMed=2434381;  
RA Leube R.E., Bosch F.X., Romano V., Zimbelmann R., Hofler H.,  
RA Franke W.W.;  
RT "Cytokeratin expression in simple epithelia. III. Detection of mRNAs  
RT encoding human cytokeratins nos. 8 and 18 in normal and tumor cells by  
RT hybridization with cDNA sequences in vitro and in situ.";  
RL Differentiation 33:69-85(1986).  
RN [8]  
RP PARTIAL PROTEIN SEQUENCE.  
RX TISSUE=Colon carcinoma;  
RX MEDLINE=97295306; PubMed=9150948;  
RA Ji H., Reid G.E., Moritz R.L., Eddes J.S., Burgess A.W., Simpson R.J.;  
RT "A two-dimensional gel database of human colon carcinoma proteins.";  
RT Electrophoresis 18:605-613(1997).  
RN [9]  
RP INTERACTION WITH PNN.  
RX PubMed=10809736; DOI=10.1074/jbc.275.20.14910;  
RA Shi J., Sugrue S.P.;  
RT "Dissection of protein linkage between keratins and pinin, a protein  
RT with dual location at desmosome-intermediate filament complex and in  
RT the nucleus.";  
RL J. Biol. Chem. 275:14910-14915(2000).  
RN [10]  
RP PHOSPHORYLATION SITE SER-73.  
RX MEDLINE=21909496; PubMed=11781324; DOI=10.1074/jbc.M111436200;  
RA He T., Stepulak A., Holmstrom T.H., Omary M.B., Eriksson J.E.;  
RT "The intermediate filament protein keratin 8 is a novel cytoplasmic  
RT substrate for c-Jun N-terminal kinase.";  
RL J. Biol. Chem. 277:10767-10774(2002).  
RN [11]  
RP PHOSPHORYLATION SITE SER-73, AND MUTAGENESIS OF LEU-71 AND SER-73.  
RX MEDLINE=21909510; PubMed=11788583; DOI=10.1074/jbc.M107623200;  
RA Ku N.O., Azhar S., Omary M.B.;  
RT "Keratin 8 phosphorylation by p38 kinase regulates cellular keratin  
RT filament reorganization: modulation by a keratin 1-like disease  
RT causing mutation.";  
RL J. Biol. Chem. 277:10775-10782(2002).  
RN [12]  
RP INTERACTION WITH HCV CORE PROTEIN.  
RX PubMed=15846844; DOI=10.1002/pmic.200401093;  
RA Kang S.-M., Shin M.-J., Kim J.-H., Oh J.-W.;  
RT "Proteomic profiling of cellular proteins interacting with the  
RT hepatitis C virus core protein.";  
RL Proteomics 5:2227-2237(2005).  
RN [13]  
RP VARIANTS CRYPTOGENIC CIRRHOSIS VAL-52; CYS-53 AND CYS-61, AND VARIANT  
RP VAL-62.  
RX PubMed=1274528; DOI=10.1073/pnas.0936165100;  
RA Ku N.-O., Darling J.M., Krams S.M., Esquivel C.O., Keefe E.B.,  
RA Sibley R.K., Lee Y.M., Wright T.L., Omary M.B.;  
RT "Keratin 8 and 18 mutations are risk factors for developing liver  
RT disease of multiple etiologies.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:6063-6068(2003).  
CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
CC Keratin 8 associates with keratin 18. Interacts with HCV core  
CC protein and PNN.  
CC -!- PTM: Phosphorylation on serine residues is enhanced during EGF  
CC stimulation and mitosis. Ser-73 phosphorylation plays an important  
CC role in keratin filament reorganization.  
CC -!- DISEASE: Defects in KRT8 are a cause of cryptogenic cirrhosis  
CC [MIM:215600].  
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
CC microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II  
CC (neutral to basic; 56-70 kDa) [K1 to K8].  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC -!- DATABASE: NAME=Human Intermediate Filament Mutation Database;  
CC WWW="http://www.interfil.org".  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial License  
CC -----  
DR EMBL; M34225; AAA35748.1; -; mRNA.  
DR EMBL; M34482; AAA35763.1; -; Genomic\_DNA.  
DR EMBL; X74929; CAA52882.1; -; mRNA.  
DR EMBL; U76549; AAB18966.1; -; mRNA.  
DR EMBL; X98614; CAA67203.1; -; mRNA.  
DR EMBL; M26512; AAA51542.1; -; mRNA.  
DR EMBL; BC000654; AAH00654.2; ALT INIT; mRNA.  
DR EMBL; X12882; CAA31376.1; -; mRNA.  
DR EMBL; X74981; CAA52916.1; -; Genomic\_DNA.  
DR PIR; A34720; A34720.  
DR HSSP; P08670; IGK7.  
DR IntAct; P05787; -.  
DR GlycoSuiteDB; P05787; -.  
DR SWISS-2DPAGE; P05787; HUMAN.  
DR Sienna-2DPAGE; P05787; -.  
DR Ensembl; ENSG00000170421; Homo sapiens.

DR HGNC: 148060; KRT8.  
 DR MIM: 148060; Gene.  
 DR MIM: 215600; Phenotype.  
 DR LinkHub; P05787; -.  
 DR GO: 0005882; C:intermediate filament; NAS.  
 DR GO: 0005198; P:structural molecule activity; NAS.  
 DR GO: 0007010; P:cytoskeleton organization and biogenesis; NAS.  
 DR InterPro: IPR001664; IF.  
 DR InterPro: IPR002957; Keratin\_I.  
 DR InterPro: IPR003054; Keratin\_II.  
 DR PANTHER: PTHR1893:SF5; Keratin\_II; 1.  
 DR Pfam: PF00038; Filament; 1.  
 DR PRINTS: PR01248; TYPE1KERATIN.  
 DR PRINTS: PR01276; TYPE2KERATIN.  
 DR PROSITE: PS00226; IF; 1.  
 DR Coiled coil; Direct protein sequencing; Disease mutation;  
 KW Intermediate filament; Keratin; Phosphorylation; Polymorphism.  
 FT INIT\_MET 0 0  
 FT CHAIN 1 482 Keratin, type II cytoskeletal 8.  
 FT REGION 1 89 Head.  
 FT REGION 90 397 Rod.  
 FT REGION 90 125 Coil 1A.  
 FT REGION 126 142 Linker 1.  
 FT REGION 143 234 Coil 1B.  
 FT REGION 235 258 Linker 12.  
 FT REGION 259 397 Coil 2.  
 FT REGION 260 381 Necessary for interaction with PNN.  
 FT REGION 398 482 Tail.  
 FT SITE 341 341 Stutter.  
 FT MOD\_RES 23 23 Phosphoserine.  
 FT MOD\_RES 73 73 Phosphoserine (by MAPK).  
 FT MOD\_RES 431 431 Phosphoserine (by CaMK2 and MAPK).  
 FT VARIANT 52 52 G -> V (in cryptogenic cirrhosis).  
 FT VARIANT 53 53 Y -> C (in cryptogenic cirrhosis).  
 FT VARIANT 61 61 G -> C (in cryptogenic cirrhosis).  
 FT VARIANT 61 61 /FTID=VAR\_023060.  
 Query Match 100.0%; Score 38; DB 1; Length 482;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YEDEINK 7  
 DB 178 YEDEINK 184  
 |||||  
 |||||

[2] NUCLEOTIDE SEQUENCE [MRNA].  
 RX MEDLINE=95042924; PubMed=7525090;  
 RA Wang X., Hsieh J.-T., Zhou H.E.;  
 RT "Cloning and characterization of a specific cytokeratin-8 cDNA from  
 RL rat prostatic epithelium.";  
 RN Zhongguo Yi Xue Ke Xue Yuan Xue Bao 16:1-7(1994).  
 [3] NUCLEOTIDE SEQUENCE [MRNA].  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RC TISSUE=Heart muscle;  
 RX PubMed=15247274; DOI=10.1074/jbc.M400128200;  
 RA Ursitti J.A., Lee P.C., Resneck W.G., McNally M.M., Bowman A.L.,  
 O'Neill A., Stone M.R., Bloch R.J.;  
 RT "Cloning and characterization of cytokeratins 8 and 19 in adult rat  
 RL striated muscle. Interaction with the dystrophin glycoprotein  
 RT complex.";  
 RL J. Biol. Chem. 279:41830-41838(2004).  
 [4] NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Placenta, and Thymus;  
 RG NIH - Mammalian Gene Collection (MGC) project;  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 [5] PROTEIN SEQUENCE OF 8-17; 23-74 AND 414-451.  
 RX MEDLINE=96220159; PubMed=8660345; DOI=10.1006/bbrc.1996.0546;  
 RA Ando S., Tokui T., Yano T., Inagaki M.;  
 RT "Keratin 8 phosphorylation in vitro by CAMP-dependent protein kinase  
 RT occurs within the amino- and carboxyl-terminal end domains.";  
 RL Biochem. Biophys. Res. Commun. 221:67-71(1996).  
 [6] PROTEIN SEQUENCE OF 264-272, AND SUBCELLULAR LOCATION.  
 RP TISSUE=Liver;  
 RX PubMed=16128803; DOI=10.1111/j.1742-4658.2005.04847.x;  
 RA Segawa M., Niino K., Mineki R., Kaga N., Murayama K., Sugimoto K.,  
 Watanabe Y., Furukawa K., Horigome T.;  
 RT "Proteome analysis of a rat liver nuclear insoluble protein fraction  
 RT and localization of a novel protein, ISP36, to compartments in the  
 RL interchromatin space.";  
 RL FEBS J. 272:4327-4338(2005).  
 CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
 CC Keratin 8 associates with keratin 18. Interacts with PNN (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear matrix.  
 CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
 CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
 CC 55 and 56-70 kilodaltons, respectively).  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: M63482; AAA19667.1; -; mRNA.  
 CC EMBL: M63482; AAA19668.1; ALT INIT; mRNA.  
 CC EMBL: S76054; -; NOT\_ANNOTATED\_CDS; mRNA.  
 CC EMBL: AY464139; AAK36875.1; -; mRNA.  
 CC EMBL: BC091106; AAH91106.1; -; mRNA.  
 CC EMBL: BC097497; AAH97497.1; -; mRNA.  
 CC HSSP: P08670; 1GK7.  
 CC Ensembl: ENSRNOG0000009779; Rattus norvegicus.  
 CC RGD: 2984; Krt2-8.  
 CC LinkHub; Q10758; -.  
 CC GO: 00045095; C:Keratin filament; IDA.  
 CC InterPro: IPR001664; IF.  
 CC InterPro: IPR002957; Keratin\_I.  
 CC InterPro: IPR03054; Keratin\_II.  
 CC PANTHER: PTHR1893:SF5; Keratin\_II; 1.  
 CC Pfam: PF00038; Filament; 1.  
 CC PRINTS: PR01248; TYPE1KERATIN.  
 CC PRINTS: PR01276; TYPE2KERATIN.  
 CC PROSITE: PS00226; IF; 1.  
 KW Coiled coil; Direct protein sequencing; Intermediate filament;  
 KW Keratin; Nuclear protein; Phosphorylation.  
 FT INIT\_MET 0 0 By similarity.  
 FT INIT\_MET 0 0

```
FT CHAIN      1  482      Keratin, type II cytoskeletal 8.  
FT REGION    1  89      /FTId=PRO_0000063742.  
FT REGION    90 397      Head.  
FT REGION    90 125      Rod.  
FT REGION    126 142      Coil 1A.  
FT REGION    143 234      Linker 1.  
FT REGION    235 258      Coil 1B.  
FT REGION    259 397      Linker 12.  
FT REGION    260 381      Coil 2.  
FT REGION    398 482      Necessary for interaction with PNN (By  
FT SITE       341 341      similarity).  
FT SITE       23 23      Tail.  
FT MOD_RES   431 431      Stutter.  
FT MOD_RES   482 AA; 53888 MW; C4DD2FE25CB2C18B CRC64;  
SQ SEQUENCE  482 AA; 53888 MW; C4DD2FE25CB2C18B CRC64;  
  
Query Match      100.0%; Score 38; DB 1; Length 482;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY      1 YEDEINK 7  
Db      178 YEDEINK 184
```

Search completed: July 5, 2006, 19:01:33  
Job time : 72.7311 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 60.8252 Seconds  
(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928a-1

Perfect score: 38

Sequence: 1 YEDEINK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_8.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*
- 10: Geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	38	100.0	7	6	AAO26381 Psoriasis
2	38	100.0	7	9	ADV04406 Leishmania
3	38	100.0	8	6	AAO26392 Psoriasis
4	38	100.0	8	9	ADV04417 Leishmania
5	38	100.0	113	6	ABG74871 Human cyt
6	38	100.0	132	8	ADR98929 Lung spec
7	38	100.0	180	8	ADN99441 Novel hum
8	38	100.0	186	4	AA16318 Peptide #
9	38	100.0	186	4	ABE35310 Peptide #
10	38	100.0	186	4	AA28813 Peptide #
11	38	100.0	186	4	ABE30140 Peptide #
12	38	100.0	186	4	ABE20752 Protein #
13	38	100.0	186	4	AA68514 Human bon
14	38	100.0	186	4	AA56141 Human bra
15	38	100.0	186	4	ABG50183 Human liv
16	38	100.0	186	4	AA04056 Peptide #
17	38	100.0	186	4	ABG09410 Novel hum
18	38	100.0	186	5	ABG38095 Human pep
19	38	100.0	233	8	ADR98930 Lung spec
20	38	100.0	273	4	ABG26376 Novel hum
21	38	100.0	275	4	ABG15282 Novel hum
22	38	100.0	278	8	ABM81219 Tumour-as
23	38	100.0	282	8	ABM80614 Tumour-as

24	38	100.0	328	7	ADC31601	Adc31601 Human nov
25	38	100.0	342	7	ADE15641	Adel15641 Human str
26	38	100.0	347	4	ABG15636	Abg15636 Novel hum
27	38	100.0	347	4	ABG09411	Abg09411 Novel hum
28	38	100.0	394	4	ABG26375	Abg26375 Novel hum
29	38	100.0	398	6	AAE38086	AAe38086 Human cyt
30	38	100.0	401	8	ABM80365	ABm80365 Tumour-as
31	38	100.0	403	8	ABM81640	ABm81640 Tumour-as
32	38	100.0	416	4	ABG23034	Abg23034 Novel hum
33	38	100.0	420	8	ADP29879	Adp29879 Human sec
34	38	100.0	429	4	ABG19547	Abg19547 Novel hum
35	38	100.0	429	4	ABG19546	Abg19546 Novel hum
36	38	100.0	476	8	ADR66033	ADR66033 Human pro
37	38	100.0	476	8	ADR66931	ADR66931 Human pro
38	38	100.0	477	4	AAU06112	AAu06112 Novel hum
39	38	100.0	480	10	AEF92273	Aef92273 Human cyt
40	38	100.0	481	6	ABU52604	ABu52604 Human NOV
41	38	100.0	482	4	ABG08132	Abg08132 Novel hum
42	38	100.0	482	6	AAE38082	AAe38082 Human cyt
43	38	100.0	482	8	ADT51418	ADt51418 Human ker
44	38	100.0	482	8	ADT51421	ADt51421 Human ker
45	38	100.0	482	8	ADT51419	ADt51419 Human ker

#### ALIGNMENTS

RESULT 1  
AAO26381  
ID AAO26381 standard; peptide; 7 AA.  
XX  
AC AAO26381;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 1.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
XX (AKIV-) AKIVA LLC.  
XX  
PI O'daly JA;  
XX  
DR WPI; 2003-018763/01.  
XX  
XX New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
PS Claim 1; Page 43; 56pp; English.  
XX  
XX The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The sequences of immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention

```

SQ Sequence 7 AA;
Query Match 100.0%; Score 38; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. NO. 2.1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 1 YEDEINK 7

RESULT 2
ADV04406
ID ADV04406 standard; peptide; 7 AA.
XX
AC ADV04406;
XX
DT 24-FEB-2005 (first entry)
XX
DE Leishmania immunotherapeutic peptide SEQ ID NO:1.
XX
KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;
KW psoriasis.
XX
OS Leishmania sp.
XX
PN US2004241168-A1.
XX
PD 02-DEC-2004.
XX
PF 09-FEB-2004; 2004US-00774928.
XX
PR 16-MAR-2001; 2001US-00809003.
PR 17-OCT-2003; 2003US-00687892.
XX
PA (ODAL/) ODALY J A.
XX
PI Odaly JA;
XX
DR WPI; 2005-011563/01.
XX
PT Inhibiting selectively T-cell rolling in human, by administering compound
PT interfering with cutaneous lymphocyte antigen-E selectin interaction and
PT leukocyte function associated antigen-1 interaction.
XX
PS Claim 9; SEQ ID NO 1; 21pp; English.
XX
CC The invention relates to a novel method for selectively inhibiting T-cell
CC rolling in a human host, comprising administering a compound that
CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E
CC selectin interaction and leukocyte function associated antigen (LFA) -
CC 1/intercellular adhesion molecule (ICAM) and very late antigen
CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound
CC includes an immunotherapeutic agent, which comprises a purified protein
CC extract that is isolated by diethylaminoethyl Sephadex chromatography of
CC a Nonidet P-40 insoluble particulate antigen fraction derived from
CC isolated killed cells of anastigotes from one or more species of the
CC Leishmania genus, where the particulate antigen fraction is solubilized
CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied
CC to diethylaminoethyl Sephadex and is eluted with a solution comprising
CC 0.1 M sodium chloride, 8 M urea and 0.025 M
CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract
CC includes polypeptides having apparent molecular weights of 73, 80 and 82
CC kDa, after total reduction and alkylation. The species is Leishmaniosis,
CC L. venezuelensis, L. braziliensis and/or L. chagasi. A compound of the
CC invention has antipsoriatic and immunostimulant activity, and inhibits T-
CC cell rolling by interfering with CLA-E selectin interaction and LFA-
CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively
CC inhibiting T-cell rolling in a human host, and for treating psoriasis.
CC The present sequence represents a Leishmania peptide used in the
CC invention.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 38; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. NO. 2.1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 1 YEDEINK 7

RESULT 3
AAO26392
ID AAO26392 standard; peptide; 8 AA.
XX
AC AAO26392;
XX
DT 30-JAN-2003 (first entry)
XX
DE Psoriasis treating immunotherapeutic peptide, SEQ ID NO 12.
XX
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;
KW gene therapy.
XX
OS Leishmania sp.
XX
PN WO200274239-A2.
XX
PD 26-SEP-2002.
XX
PF 04-MAR-2002; 2002WO-US006496.
XX
PR 16-MAR-2001; 2001US-00809003.
XX
PA (AKIV-) AKIVA LLC.
XX
PI O'daly JA;
XX
DR WPI; 2003-018763/01.
XX
PT New polypeptide and immunogenic variants comprising amino acid sequences
PT of particulate antigens, useful for the treatment and clinical remission
PT of psoriasis.
XX
PS Claim 1; Page 43; 56pp; English.
XX
CC The invention relates to a polypeptide comprising an isolated amino acid
CC sequence or immunogenic variants selected from any of 14 fully defined
CC sequences of 7-16 amino acids, given in the specification. The
CC immunotherapeutic agents and a pharmaceutical compositions comprising
CC polynucleotides and vectors of the invention are useful for the treatment
CC and clinical remission of psoriasis. The isolated nucleic acids are
CC useful as probes. The sequences of the invention can be used in the
CC treatment of disorders by gene therapy. This sequence represents one of
CC the 14 immunotherapeutic peptides of the invention
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 6; Length 8;
Best Local Similarity 100.0%; Pred. NO. 2.1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 2 YEDEINK 8

RESULT 4
ADV04417
ID ADV04417 standard; peptide; 8 AA.
XX
AC ADV04417;
XX

```



DT 24-FEB-2005 (first entry)  
XX Leishmania immunotherapeutic peptide SEQ ID NO:12.  
DE T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;  
XX psoriasis.  
KW Leishmania sp.  
XX US2004241168-A1.  
XX 02-DEC-2004.  
XX 09-FEB-2004; 2004US-00774928.  
XX 16-MAR-2001; 2001US-00809003.  
XX 17-OCT-2003; 2003US-00687892.  
XX (ODAL/) ODALY J A.  
XX Odaly JA;  
PI WPI; 2005-011563/01.  
XX Inhibiting selectively T-cell rolling in human, by administering compound  
XX interfering with cutaneous lymphocyte antigen-E selectin interaction and  
XX leukocyte function associated antigen-1 interaction.  
XX Claim 18; SEQ ID NO 12; 21pp; English.  
XX The invention relates to a novel method for selectively inhibiting T-cell  
XX rolling in a human host, comprising administering a compound that  
XX selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
XX selectin interaction and leukocyte function associated antigen (LFA)-  
XX 1/intercellular adhesion molecule (ICAM) and very late antigen  
XX (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
XX includes an immunotherapeutic agent, which comprises a purified protein  
XX extract that is isolated by diethylaminoethyl Sephadex chromatography of  
XX a Nonidet P-40 insoluble particulate antigen fraction derived from  
XX isolated killed cells of amastigotes from one or more species of the  
XX Leishmania genus, where the particulate antigen fraction is solubilized  
XX with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
XX to diethylaminoethyl Sephadex and is eluted with a solution comprising  
XX 0.1 M sodium chloride, 8 M urea and 0.025 M  
XX Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract  
XX includes polypeptides having apparent molecular weights of 73, 80 and 82  
XX kDa, after total reduction and alkylation. The species is L.amazonensis ,  
XX L.venezuelensis , L.brasiliensis and/or L.chagasi. A compound of the  
XX invention has antipsoriatic and immunostimulant activity, and inhibits T-  
XX cell rolling by interfering with CLA-E selectin interaction and LFA-  
XX 1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
XX inhibiting T-cell rolling in a human host, and for treating psoriasis.  
XX The present sequence represents a Leishmania peptide used in the  
XX invention.  
XX Sequence 8 AA;  
SQ Query Match 100.0%; Score 38; DB 9; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 YEDEINK 7  
DB 2 YEDEINK 8  
RESULT 5  
ABG74871  
ID ABG74871 standard; protein; 113 AA.  
XX AC ABG74871;  
XX DT 19-JUN-2003 (first entry)  
XX Human cytokерatin-1 protein fragment.  
XX Cytokeratin-1; inflammation; infection; soluble; sepsis; veterinary;  
KW sepsis-like systemic infection; human.  
XX Homo sapiens.  
XX WO2003002600-A1.  
XX 09-JAN-2003.  
XX 12-JUN-2002; 2002WO-EP006473.  
XX 27-JUN-2001; 2001DE-01030985.  
XX (BRAH-) BRAHMS AG.  
XX Bergmann A, Struck J, Uehlein M;  
XX WPI; 2003-201491/19.  
XX New use of soluble cytokерatin-1 fragments from body fluid are tissue as  
XX marker peptides or to provide therapeutics for the diagnosis, prognosis  
XX and treatment of inflammatory disease and infection, including systemic  
XX inflammation.  
XX Claim 4; Page 41; 51pp; German.  
XX This invention describes a novel use of soluble cytokерatin-1 fragments  
XX from body fluid or tissue as marker peptides for the detection, prognosis  
XX and control of inflammation and infection or as a therapeutic target for  
XX inflammation and infection. The invention also describes a method of  
XX producing a soluble cytokерatin-1 fragment in rich or pure form from a  
XX human material rich in cytokерatin-1 by enzymatically splitting the  
XX cytokерatin-1 with endoprotease to obtain the hydrolysis products and  
XX purifying the resulting fragments. The novel cytokерatin fragments are  
XX useful for the differential diagnosis, early recognition and prognosis,  
XX evaluation of severity of sepsis and heavy infection, particularly sepsis  
XX -like systemic infection. The fragments are used in human and veterinary  
XX therapy, particularly the fragment or it's specific antibody is used to  
XX treat inflammatory disease or infection. This sequence represents the  
XX human cytokерatin-1 protein fragment described in the disclosure of the  
XX invention  
SQ Sequence 113 AA;  
Query Match 100.0%; Score 38; DB 6; Length 113;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 YEDEINK 7  
DB 85 YEDEINK 91  
RESULT 6  
ADR98929  
ID ADR98929 standard; protein; 132 AA.  
XX AC ADR98929;  
XX 18-NOV-2004 (first entry)  
XX Lung specific gene splice variant encoded protein #101.  
XX cytotactic; gene therapy; vaccine; lung; diagnosis; cancer;  
KW non-cancerous lung disease; lung tissue; antagonist; gene therapy;  
KW transgenic animal; splice variant.  
XX OS Homo sapiens.  
XX WO2004074430-A2.  
PN

```
XX
PD 02-SEP-2004.
XX
PF 08-DEC-2003; 2003WO-US038896.
XX
PR 06-DEC-2002; 2002US-0431307P.
PR 06-DEC-2002; 2002US-0431510P.
PR 06-DEC-2002; 2002US-0431516P.
XX
PA (DIAD-) DIADEXUS INC.
XX
XX Macina RA, Turner LR, Sun Y, Liu S;
XX
XX WPI, 2004-635553/61.
DR N-FSDB; ADR98791.
XX
XX New isolated human lung specific nucleic acid molecule, useful for
PT identifying, diagnosing, monitoring, staging, imaging and treating lung
PT cancer and non-cancerous diseases of the lung.
XX
PS Claim 1; SEQ ID NO 214; 542pp; English.
XX
CC The invention relates to a new isolated lung specific nucleic acid
CC molecule (I) comprising any of 113 fully defined nucleotide sequences
CC given in the specification, their encoded protein sequences, sequences
CC selectively hybridizing to the nucleotide sequences or a sequence having
CC at least 60% identity to the nucleotide sequences. The methods and
CC compositions of the present invention are useful for identifying,
CC diagnosing, monitoring, staging, imaging and treating lung cancer and non
CC cancerous diseases of the lung. They are also used for identifying lung
CC tissue, monitoring and identifying and/or designing antagonists of the
CC polypeptide of the invention, gene therapy, production of transgenic
CC animals and production of engineered lung tissue for treatment and
CC research. Lung specific genes (LSEs) were identified by a systematic
CC analysis of gene expression data in the LIFESEQ Gold database using the
CC data mining software package candidate lead automatic search program
CC (CLASP). Genes were grouped into gene bins where each bin is a cluster of
CC sequences grouped together where they share a common contig.
CC Differentially expressed tissue-specific genes were selected based on the
CC percentage level in the targeted tissue versus all the other tissues. The
CC expression levels for each gene in libraries of normal tissues or non-
CC tumour tissues from cancer patients were compared with the expression
CC levels in tissue libraries associated with tumour or disease. This
CC sequence represents a protein of the invention.
XX
SQ Sequence 132 AA;
Query Match 100.0%; Score 38; DB 8; Length 132;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YEDEINK 7
DB 47 YEDEINK 53
RESULT 7
ADN99441
ID ADN99441 standard; protein; 180 AA.
XX
AC ADN99441;
XX
DT 29-JUL-2004 (first entry)
XX
DE Novel human protein sequence #257.
XX
KW anti-inflammatory; dermatological; neuroprotective; immunomodulator;
KW antibacterial; virucide; antiparasitic; cytostatic; gene therapy;
KW vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
KW early aging; hormonal imbalance; ischemic heart disease;
KW ulcerative colitis.
XX
XX Homo sapiens.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
```

PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;  
PI Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;  
PI Wong JGP, Wu G, Zhang H, Zeng C;  
XX  
DR WPI: 2004-365511/34.  
XX N-PSDB; ADN98657.  
XX  
PT New nucleic acid molecules, useful in preparing a composition for  
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral  
PT disorders, cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic  
PT ulcerative colitis.  
XX  
XX Claim 14; SEQ ID NO 1041; 532pp; English.  
PS  
XX  
XX The invention relates to a nucleic acid molecule comprising a  
CC polynucleotide sequence or its complement that encodes a polypeptide. The  
CC nucleic acid is useful in preparing a composition for treating or  
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,  
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic  
CC heart disease or ulcerative colitis. This sequence corresponds to a  
CC protein of the invention.  
XX  
XX Sequence 180 AA;  
SQ  
Query Match 100.0%; Score 38; DB 8; Length 180;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YEDEINK 7  
DB 41 YEDEINK 47  
RESULT 8  
AAM16318  
ID AAM16318 standard; protein; 186 AA.  
XX  
AC AAM16318;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #2752 encoded by probe for measuring cervical gene expression.  
XX  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.  
XX  
OS Homo sapiens.  
XX  
XX WO200157278-A2.  
PN  
PD 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000670.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI: 2001-488901/53.  
DR  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.  
PT  
XX  
XX Claim 27; SEQ ID NO 21144; 487pp; English.  
PS  
XX

CC The present invention relates to human single exon nucleic acid probes  
CC (SENP: see AA110068-AA128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 186 AA;  
SQ  
Query Match 100.0%; Score 38; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YEDEINK 7  
DB 174 YEDEINK 180  
RESULT 9  
ABB35310  
ID ABB35310 standard; peptide; 186 AA.  
XX  
AC ABB35310;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #2816 encoded by human foetal liver single exon probe.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
OS Homo sapiens.  
XX  
XX WO200157277-A2.  
PN  
PD 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000669.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI: 2001-483447/52.  
DR  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human foetal liver.  
PT  
XX  
XX Claim 27; SEQ ID NO 27945; 639pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human foetal liver. The  
XX present sequence is a peptide encoded by a single exon nucleic acid probe  
XX of the invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 186 AA;  
SQ  
Query Match 100.0%; Score 38; DB 4; Length 186;

```

Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 174 YEDEINK 180

RESULT 10
AAM28813
ID AAM28813 standard; protein; 186 AA.
XX
AC AAM28813;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #2850 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human placenta.
XX
Claim 27; SEQ ID NO 29082; 654pp; English.
XX
The present invention relates to single exon nucleic acid probes (SENP:
see AAI31315-AAI57546). The present sequence is a peptide encoded by one
such probe. The probes are useful for producing a microarray for
predicting, measuring and displaying gene expression in samples derived
from human placenta. The probes are useful for antenatal diagnosis of
human genetic disorders
XX
SQ Sequence 186 AA;
Query Match 100.0%; Score 38; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 174 YEDEINK 180

RESULT 11
ABB30140
ID ABB30140 standard; peptide; 186 AA.
XX
AC ABB30140;
XX
DT 01-FEB-2002 (first entry)
DE Protein #2751 encoded by probe for measuring heart cell gene expression.

```

```

XX Peptide #2791 encoded by breast cell single exon nucleic acid probe.
DE
XX Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
New spatially-addressable set of single exon nucleic acid probes, useful
for measuring gene expression in sample derived from human breast,
comprises number of single exon nucleic acid probes.
XX
Claim 27; SEQ ID NO 13108; 327pp + Sequence Listing; English.
XX
The invention relates to a spatially-addressable set of single exon
nucleic acid probes for measuring gene expression in a sample derived
from human breast and Br 474 cells. The method involves contacting the
probes with a collection of detectably labelled nucleic acids derived
from mRNA of human breast, and then measuring the label bound to each
probe of the microarray. The probes are useful for verifying the
expression of regions of genomic DNA predicted to encode proteins. They
are useful for gene discovery, and for determining predisposition and/or
prognosing breast disease. Gene expression analysis is useful for
assessing the toxicity of chemical agents on cells. The microarray of
this invention presents a far greater diversity of probes for measuring
gene expression, with far less bias than expressed sequence tag
microarrays. The method is suitable for rapid production of functional
information from genomic sequence. The present sequence is a peptide
encoded by a single exon nucleic acid probe of the invention. Note: The
sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 186 AA;
Query Match 100.0%; Score 38; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 174 YEDEINK 180

RESULT 12
ABB20752
ID ABB20752 standard; protein; 186 AA.
XX
AC ABB20752;
XX
DT 23-JAN-2002 (first entry)
DE Protein #2751 encoded by probe for measuring heart cell gene expression.

```

```
XX KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX Homo sapiens.
OS WO200157274-A2.
PN 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488999/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX Claim 15; SEQ ID NO 22522; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 186 AA;
Query Match 100.0%; Score 38; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YEDEINK 7
Db 174 YEDEINK 180
RESULT 13
AAM68514
ID AAM68514 standard; protein; 186 AA.
XX AC AAM68514;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 28820.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488999/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX Example 4; SEQ ID NO 28820; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX SQ Sequence 186 AA;
Query Match 100.0%; Score 38; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YEDEINK 7
Db 174 YEDEINK 180
RESULT 14
AAM56141
ID AAM56141 standard; protein; 186 AA.
XX AC AAM56141;
XX 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 28246.
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX Homo sapiens.
XX WO200157275-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000667.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
```

XX WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains.  
XX Example 4; SEQ ID NO 28246; 650pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention  
XX  
SQ Sequence 186 AA;  
Query Match 100.0%; Score 38; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 YEDEINK 7  
Db 174 YEDEINK 180  
RESULT 15  
ABG50183  
ID ABG50183 standard; peptide; 186 AA.  
AC ABG50183;  
XX  
DT 25-FEB-2003 (first entry)  
DE Human liver peptide, SEQ ID NO 28831.  
XX  
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157273-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US0000664.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488898/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human adult liver.  
XX  
PS Claim 27; SEQ ID NO 28831; 658pp; English.  
XX  
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression

CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABG47348-ABG5930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 186 AA;  
Query Match 100.0%; Score 38; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 YEDEINK 7  
Db 174 YEDEINK 180  
Search completed: July 5, 2006, 18:43:53  
Job time : 60.8252 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 51.2427 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-1

Perfect score: 38

Sequence: 1 YEDEINK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	7	5	US-10-774-928-1
2	38	100.0	8	5	US-10-774-928-12
3	38	100.0	113	5	US-10-480-848A-5
4	38	100.0	186	3	US-09-864-761-36050
5	38	100.0	186	5	US-10-450-763-39769
6	38	100.0	273	5	US-10-450-763-56735
7	38	100.0	275	5	US-10-450-763-45641
8	38	100.0	347	5	US-10-450-763-39770
9	38	100.0	347	5	US-10-450-763-45995
10	38	100.0	394	5	US-10-450-763-56734
11	38	100.0	398	5	US-10-883-020-5
12	38	100.0	416	5	US-10-450-763-53393
13	38	100.0	422	3	US-09-779-307-18
14	38	100.0	422	3	US-09-779-307-20
15	38	100.0	429	5	US-10-450-763-49905
16	38	100.0	429	5	US-10-450-763-49906
17	38	100.0	430	3	US-09-779-307-19
18	38	100.0	476	3	US-09-779-307-16
19	38	100.0	477	3	US-09-779-307-4
20	38	100.0	481	4	US-10-080-334-54
21	38	100.0	481	6	US-11-156-300-24
22	38	100.0	482	3	US-09-779-307-17
23	38	100.0	482	4	US-10-080-334-187
24	38	100.0	482	5	US-10-883-020-1
25	38	100.0	482	5	US-10-450-763-38491
26	38	100.0	483	3	US-09-919-497-79
27	38	100.0	483	3	US-09-974-298-41

```

28 38 100.0 483 3 US-09-981-353-158 Sequence 158, App
29 38 100.0 483 4 US-10-177-293-258 Sequence 258, App
30 38 100.0 483 4 US-10-080-334-183 Sequence 183, App
31 38 100.0 483 4 US-10-080-334-184 Sequence 184, App
32 38 100.0 483 4 US-10-080-334-185 Sequence 185, App
33 38 100.0 483 4 US-10-080-334-186 Sequence 186, App
34 38 100.0 483 4 US-10-734-564-116 Sequence 116, App
35 38 100.0 483 6 US-11-037-713-9 Sequence 2, Appli
36 38 100.0 498 4 US-10-605-708A-2 Sequence 2, Appli
37 38 100.0 500 5 US-10-450-763-39773 Sequence 39773, A
38 38 100.0 513 5 US-10-450-763-38492 Sequence 38492, A
39 38 100.0 521 3 US-09-976-782-8 Sequence 8, Appli
40 38 100.0 523 3 US-09-976-782-6 Sequence 6, Appli
41 38 100.0 541 5 US-10-450-763-55674 Sequence 55674, A
42 38 100.0 551 3 US-09-976-782-89 Sequence 89, Appl
43 38 100.0 551 3 US-09-976-782-90 Sequence 90, Appl
44 38 100.0 551 4 US-10-015-115-127 Sequence 127, App
45 38 100.0 551 4 US-10-231-913-116 Sequence 116, App

```

## ALIGNMENTS

```

RESULT 1
US-10-774-928-1
; Sequence 1, Application US/10774928
; Publication No. US20040241168A1
; GENERAL INFORMATION:
; APPLICANT: O'Daly, Jose
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission
; TITLE OF INVENTION: Psoriasis
; FILE REFERENCE: 302178
; CURRENT APPLICATION NUMBER: US/10/774,928
; CURRENT FILING DATE: 2004-02-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Leishmania
US-10-774-928-1

```

```

Query Match      100.0%; Score 38; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.9e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 YEDEINK 7
Db 1 YEDEINK 7

```

```

RESULT 2
US-10-774-928-12
; Sequence 12, Application US/10774928
; Publication No. US20040241168A1
; GENERAL INFORMATION:
; APPLICANT: O'Daly, Jose
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission
; TITLE OF INVENTION: Psoriasis
; FILE REFERENCE: 302178
; CURRENT APPLICATION NUMBER: US/10/774,928
; CURRENT FILING DATE: 2004-02-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Leishmania
US-10-774-928-12

```

```

Query Match      100.0%; Score 38; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.9e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 YEDEINK 7
      |||||
Db      2 YEDEINK 8

RESULT 3
US-10-480-848A-5
; Sequence 5, Application US/10480848A
; Publication No. US20040219597A1
; GENERAL INFORMATION:
; APPLICANT: B.R.A.H.M.S Aktiengesellschaft
; TITLE OF INVENTION: Use of soluble cytokestatin 1 fragments in diagnosis and
; FILE REFERENCE: 3537PCTNAT
; CURRENT APPLICATION NUMBER: US/10/480.848A
; CURRENT FILING DATE: 2003-12-15
; PRIOR FILING DATE: 2001-06-27
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-480-848A-5

Query Match      100.0%; Score 38; DB 5; Length 113;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINK 7
      |||||
Db      85 YEDEINK 91

RESULT 4
US-09-864-761-36050
; Sequence 36050, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36050
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011309.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 72
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 59
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 44
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5e+02
; OTHER INFORMATION: EST HUMAN HIT: BE870853.1, EVALUE 9.00e-68
; OTHER INFORMATION: SWISSPROT HIT: P05787, EVALUE 1.00e-68
US-09-864-761-36050

```

```

Query Match      100.0%; Score 38; DB 3; Length 186;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 YEDEINK 7
      |||||
Db      174 YEDEINK 180

```

```

RESULT 5
US-10-450-763-39769
; Sequence 39769, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 39769
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (124)..(171)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226D, p-value=7.000e-38, raw score of 19.10
; FEATURE:
; NAME/KEY: DOMAIN

```



```
; LOCATION: (1)..(169)
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,
; OTHER INFORMATION: accession name filament, E-value=4.3e-71, Pfam score of 249.6
US-10-450-763-39769

Query Match      100.0%; Score 38; DB 5; Length 186;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 12 YEDEINK 18

RESULT 6
US-10-450-763-56735
; Sequence 56735, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 56735
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (177)..(225)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226B, p-value=3.348e-35, raw score of 23.86
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (78)..(271)
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,
; OTHER INFORMATION: accession name filament, E-value=8.6e-101, Pfam score of 348.3
US-10-450-763-56735

Query Match      100.0%; Score 38; DB 5; Length 273;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 167 YEDEINK 173

RESULT 7
US-10-450-763-45641
; Sequence 45641, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
```

```
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 45641
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (179)..(227)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226B, p-value=3.348e-35, raw score of 23.86
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (87)..(273)
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,
; OTHER INFORMATION: accession name filament, E-value=3.2e-89, Pfam score of 309.8
US-10-450-763-45641

Query Match      100.0%; Score 38; DB 5; Length 275;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 169 YEDEINK 175

RESULT 8
US-10-450-763-39770
; Sequence 39770, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 39770
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (180)..(228)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226B, p-value=1.643e-31, raw score of 23.86
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (77)..(342)
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,
; OTHER INFORMATION: accession name filament, E-value=1.1e-78, Pfam score of 274.8
US-10-450-763-39770

Query Match      100.0%; Score 38; DB 5; Length 347;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7
Db 170 YEDEINK 176

RESULT 9
US-10-450-763-45995
; Sequence 45995, Application US/10450763
```

```
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 45995
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (180)..(228)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226B, p-value=1.643e-31, raw score of 23.86
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (77)..(342)
; OTHER INFORMATION: Intermediate filament proteins domain identified by PFam,
; OTHER INFORMATION: accession name filament, E-value=1.1e-78, PFam score of 274.8
US-10-450-763-45995

Query Match          100.0%; Score 38; DB 5; Length 347;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
    |||||
Db 170 YEDEINK 176

RESULT 10
US-10-450-763-56734
; Sequence 56734, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 56734
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (164)..(212)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226B, p-value=4.656e-29, raw score of 23.86
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (80)..(312)
; OTHER INFORMATION: Intermediate filament proteins domain identified by PFam,
; OTHER INFORMATION: accession name filament, E-value=4e-88, PFam score of 306.2
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (1)..(394)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-56734

Query Match          100.0%; Score 38; DB 5; Length 394;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
    |||||
Db 154 YEDEINK 160

RESULT 11
US-10-883-020-5
; Sequence 5, Application US/10883020
; Publication No. US20050048070A1
; GENERAL INFORMATION:
; APPLICANT: Ditzel, Henrik
; APPLICANT: Jensenius, Jens
; APPLICANT: The Scripps Research Institute
; TITLE OF INVENTION: Cancer-Associated Epitope
; FILE REFERENCE: 1361.017US1
; CURRENT APPLICATION NUMBER: US/10/883,020
; CURRENT FILING DATE: 2004-07-01
; PRIOR APPLICATION NUMBER: PCT/US03/00297
; PRIOR FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US 60/345,208
; PRIOR FILING DATE: 2002-01-03
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-883-020-5

Query Match          100.0%; Score 38; DB 5; Length 398;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
    |||||
Db 94 YEDEINK 100

RESULT 12
US-10-450-763-53393
; Sequence 53393, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 53393
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (122)..(137)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
```

OTHER INFORMATION: accession number BL00226A, p-value=4.600e-14, raw score of 12.77

FEATURE: NAME/KEY: DOMAIN  
LOCATION: (125)..(346)  
OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam.  
OTHER INFORMATION: accession name filament, E-value=2e-54, Pfam score of 194.2  
FEATURE: NAME/KEY: misc\_feature  
LOCATION: (1)..(416)  
OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-53393

Query Match 100.0%; Score 38; DB 5; Length 416;

Best Local Similarity 100.0%; Pred. No. 72; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7

Db 212 YEDEINK 218

RESULT 13

US-09-779-307-18

Sequence 18, Application US/09779307

Patent No. US20020137675A1

GENERAL INFORMATION:

APPLICANT: Taupier Jr., Raymond

APPLICANT: Majumder, Kumud

APPLICANT: Vernet, Corine

APPLICANT: Pravaga, Sudhirdas

TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby

FILE REFERENCE: 15966-662 US

CURRENT APPLICATION NUMBER: US/09/779,307

CURRENT FILING DATE: 2001-02-07

PRIOR APPLICATION NUMBER: 60/180,880

PRIOR FILING DATE: 2000-02-08

PRIOR APPLICATION NUMBER: 60/181,044

PRIOR FILING DATE: 2000-02-08

PRIOR APPLICATION NUMBER: 60/181,656

PRIOR FILING DATE: 2000-02-10

PRIOR APPLICATION NUMBER: 60/182,795

PRIOR FILING DATE: 2000-02-15

NUMBER OF SEQ ID NOS: 43

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 18

LENGTH: 422

TYPE: PRT

ORGANISM: Homo sapiens

US-09-779-307-18

Query Match

Best Local Similarity 100.0%; Score 38; DB 3; Length 422;

Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7

Db 178 YEDEINK 184

RESULT 14

US-09-779-307-20

Sequence 20, Application US/09779307

Patent No. US20020137675A1

GENERAL INFORMATION:

APPLICANT: Taupier Jr., Raymond

APPLICANT: Majumder, Kumud

APPLICANT: Vernet, Corine

APPLICANT: Pravaga, Sudhirdas

TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby

FILE REFERENCE: 15966-662 US

CURRENT APPLICATION NUMBER: US/09/779,307

CURRENT FILING DATE: 2001-02-07

PRIOR APPLICATION NUMBER: 60/180,880

PRIOR FILING DATE: 2000-02-08  
PRIOR APPLICATION NUMBER: 60/181,044  
PRIOR FILING DATE: 2000-02-08  
PRIOR APPLICATION NUMBER: 60/181,656  
PRIOR FILING DATE: 2000-02-10  
PRIOR APPLICATION NUMBER: 60/182,795  
PRIOR FILING DATE: 2000-02-15  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 422  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
US-09-779-307-20

Query Match 100.0%; Score 38; DB 3; Length 422;

Best Local Similarity 100.0%; Pred. No. 73; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7

Db 178 YEDEINK 184

RESULT 15

US-10-450-763-49905

Sequence 49905, Application US/10450763

Publication No. US20050196754A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc

TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 790CIP3/US

CURRENT APPLICATION NUMBER: US/10/450,763

CURRENT FILING DATE: 2003-06-11

PRIOR APPLICATION NUMBER: PCT/US01/08631

PRIOR FILING DATE: 2001-03-30

PRIOR APPLICATION NUMBER: 09/540,217

PRIOR FILING DATE: 2000-03-31

PRIOR APPLICATION NUMBER: 09/649,167

PRIOR FILING DATE: 2000-08-23

NUMBER OF SEQ ID NOS: 60736

SOFTWARE: Custom

SEQ ID NO 49905

LENGTH: 429

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: DOMAIN

LOCATION: (183)..(231)

OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,

OTHER INFORMATION: accession number BL00226B, p-value=2.714e-23, raw score of 23.86

FEATURE:

NAME/KEY: DOMAIN

LOCATION: (107)..(348)

OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,

OTHER INFORMATION: accession name filament, E-value=4.2e-90, Pfam score of 312.7

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)..(429)

OTHER INFORMATION: Xaa = X or \* as defined in Table 2

US-10-450-763-49905

Query Match

Best Local Similarity 100.0%; Score 38; DB 5; Length 429;

Mismatches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7

Db 173 YEDEINK 179

Search completed: July 5, 2006, 20:52:00

Job time : 52.2427 secs

---

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 4.00971 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-1

Perfect score: 38  
Sequence: 1 YEDEINK 7

Scoring table: BLOSUM62

Gapop 10:0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	483	7	US-11-105-233-180
2	35	92.1	469	6	US-10-505-928-457
3	34	89.5	140	6	US-10-953-349-1458
4	34	89.5	147	6	US-10-953-349-1457
5	34	89.5	164	6	US-10-953-349-1456
6	34	89.5	940	6	US-10-449-902-45238
7	34	89.5	940	6	US-10-449-902-45669
8	33	86.8	242	6	US-10-449-902-48212
9	33	86.8	256	7	US-11-174-307B-4514
10	33	86.8	758	6	US-10-449-902-51086
11	33	86.8	799	6	US-10-449-902-47602
12	33	86.8	964	7	US-11-289-102-392
13	32	84.2	33	7	US-11-134-871-915
14	32	84.2	225	6	US-10-471-571A-5244
15	31	81.6	107	6	US-10-953-349-13561
16	31	81.6	158	6	US-10-953-349-13560
17	31	81.6	247	6	US-10-449-902-54216
18	31	81.6	568	6	US-10-449-902-46926
19	31	81.6	568	6	US-10-449-902-47019
20	31	81.6	591	6	US-10-953-349-1322
21	31	81.6	705	6	US-10-953-349-1321
22	31	81.6	713	6	US-10-953-349-1320
23	31	81.6	714	6	US-10-449-902-54951
24	31	81.6	723	6	US-10-449-902-53379
25	30	78.9	196	6	US-10-953-349-19059

26	30	78.9	228	6	US-10-449-902-50750	Sequence 50750, A
27	30	78.9	238	6	US-10-953-349-19058	Sequence 19058, A
28	30	78.9	350	6	US-10-449-902-55155	Sequence 55155, A
29	30	78.9	586	6	US-10-505-928-314	Sequence 314, App
30	29	76.3	158	6	US-10-953-349-13251	Sequence 13251, A
31	29	76.3	183	6	US-10-953-349-13250	Sequence 13250, A
32	29	76.3	246	6	US-10-449-902-35293	Sequence 35293, A
33	29	76.3	256	6	US-10-449-902-37092	Sequence 37092, A
34	29	76.3	407	6	US-10-449-902-51903	Sequence 51903, A
35	29	76.3	471	6	US-10-449-902-36733	Sequence 36733, A
36	29	76.3	471	6	US-10-449-902-49633	Sequence 49633, A
37	29	76.3	471	6	US-10-449-902-54344	Sequence 54344, A
38	29	76.3	531	6	US-10-953-349-27961	Sequence 27961, A
39	29	76.3	533	6	US-10-953-349-27960	Sequence 27960, A
40	29	76.3	557	6	US-10-953-349-27959	Sequence 27959, A
41	29	76.3	600	6	US-10-449-902-38563	Sequence 38563, A
42	29	76.3	600	6	US-10-449-902-54061	Sequence 54061, A
43	28	73.7	68	6	US-10-471-571A-5136	Sequence 5136, Ap
44	28	73.7	107	6	US-10-471-571A-1754	Sequence 1754, Ap
45	28	73.7	219	6	US-10-953-349-2869	Sequence 2869, Ap

## ALIGNMENTS

RESULT 1  
US-11-105-233-180  
; Sequence 180, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Thialgalingam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; FILE OF INVENTION: Tumors  
; FILE REFERENCE: 1657/2001  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; CURRENT FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-105-233-180

Query Match 100.0%; Score 38; DB 7; Length 483;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
DB 179 YEDEINK 185

RESULT 2  
US-10-505-928-457  
; Sequence 457, Application US/10505928  
; Publication No. US20060088532A1  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig Institute for Cancer Research et al.  
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES  
; FILE REFERENCE: 28967/39178  
; CURRENT APPLICATION NUMBER: US/10/505,928  
; CURRENT FILING DATE: 2004-08-27  
; PRIOR APPLICATION NUMBER: US 60/363,019  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: PatentIn 3.2  
; SEQ ID NO 457  
; LENGTH: 469  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-505-928-457

```
Query Match      92.1%; Score 35; DB 6; Length 469;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 180 YEDEINK 186

RESULT 3
US-10-953-349-1458
; Sequence 1458, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1458
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1458

Query Match      89.5%; Score 34; DB 6; Length 140;
Best Local Similarity 85.7%; Pred. No. 5.6;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 56 FEDEINK 62

RESULT 4
US-10-953-349-1457
; Sequence 1457, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1457
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1457

Query Match      89.5%; Score 34; DB 6; Length 147;
Best Local Similarity 85.7%; Pred. No. 5.9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 63 FEDEINK 69

RESULT 5
US-10-953-349-1456
; Sequence 1456, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1456
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1456

Query Match      89.5%; Score 34; DB 6; Length 164;
Best Local Similarity 85.7%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 80 FEDEINK 86

RESULT 6
US-10-449-902-45238
; Sequence 45238, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45238
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-45238

Query Match      89.5%; Score 34; DB 6; Length 940;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 413 YEDEINK 419

RESULT 7
US-10-449-902-45669
; Sequence 45669, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
```

; SEQ ID NO 45669  
; LENGTH: 940  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-45669

Query Match 89.5%; Score 34; DB 6; Length 940;  
Best Local Similarity 85.7%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 YEDEINK 7  
|||||  
DB 413 YEDEINE 419

RESULT 8  
US-10-449-902-46212  
; Sequence 46212, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46212  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-46212

Query Match 86.8%; Score 33; DB 6; Length 242;  
Best Local Similarity 85.7%; Pred. No. 16;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YEDEINK 7  
|||||  
DB 84 YEDEANK 90

RESULT 9  
US-11-174-307B-4514  
; Sequence 4514, Application US/11174307B  
; Publication No. US20060143729A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai  
; APPLICANT: BROVER, Vyacheslav  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY  
; TITLE OF INVENTION: USEFUL FOR MODIFYING PLANT CHARACTERISTICS  
; FILE REFERENCE: 2750-1601PUS2  
; CURRENT APPLICATION NUMBER: US/11/174,307B  
; CURRENT FILING DATE: 2005-06-30  
; PRIOR APPLICATION NUMBER: 60/583,671  
; PRIOR FILING DATE: 2004-06-30  
; PRIOR APPLICATION NUMBER: 60/583,781  
; PRIOR FILING DATE: 2004-06-30  
; PRIOR APPLICATION NUMBER: 60/583,651  
; PRIOR FILING DATE: 2004-06-30  
; NUMBER OF SEQ ID NOS: 5544  
; SEQ ID NO 4514  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:

; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 50947575; NR Description: thylakoid lumen protein,  
; OTHER INFORMATION: chloroplast precursor-like [Oryza sativa (japonica cultivar-group)]  
; OTHER INFORMATION: >gi|42408803|dbj|BAD10064.1| thylakoid lumen protein, chloroplast  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 21537211; NR Description: thylakoid lumen protein,  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 32815945; NR Description: Atlg76450 [Arabidopsis  
; OTHER INFORMATION: thaliana] >gi|18203439|sp|Q9S720|THL1\_ARATH Unknown thylakoid lumen  
; OTHER INFORMATION: protein, chloroplast precursor >gi|18411110|ref|NP\_565131.1|  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 61826057; NR Description: PREDICTED: similar to  
; OTHER INFORMATION: serine/arginine repetitive matrix 1, partial [Bos taurus]  
; OTHER INFORMATION: gi|61818041|ref|XP\_593433.1| PREDICTED: similar to serine/arginin  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 7949115; NR Description: Ser/Arg-related nuclear  
; OTHER INFORMATION: matrix protein; plenty-of-prolines-101; serine/arginine repetitive  
; OTHER INFORMATION: matrix protein 1 [Mus musculus] >gi|3153821|gb|AAC17422.1|  
US-11-174-307B-4514

Query Match 86.8%; Score 33; DB 7; Length 256;  
Best Local Similarity 85.7%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YEDEINK 7  
|||||  
DB 98 YEDEANK 104

RESULT 10  
US-10-449-902-51086  
; Sequence 51086, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51086  
; LENGTH: 758  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-51086

Query Match 86.8%; Score 33; DB 6; Length 758;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YEDEIN 6  
|||||  
DB 230 YEDEIN 235

RESULT 11

```
US-10-449-902-47602
; Sequence 47602, Application US/10449902
; Publication No. US20060123503A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0203Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47602
; LENGTH: 799
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-47602

Query Match      86.8%; Score 33; DB 6; Length 799;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEIN 6
| | | | |
Db 349 YEDEIN 354

RESULT 12
US-11-289-102-392
; Sequence 392, Application US/11289102
; Publication No. US20060121511A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Hyerim
; APPLICANT: Shaw, Peter M.
; APPLICANT: Clark, Edwin
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO
; TITLE OF INVENTION: MICROTUBULE-STABILIZING AGENTS
; FILE REFERENCE: 10338 NP
; CURRENT APPLICATION NUMBER: US/11/289,102
; CURRENT FILING DATE: 2005-11-29
; PRIOR APPLICATION NUMBER: US 60/631,993
; PRIOR FILING DATE: 2004-11-30
; NUMBER OF SEQ ID NOS: 395
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 392
; LENGTH: 964
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-289-102-392

Query Match      86.8%; Score 33; DB 7; Length 964;
Best Local Similarity 71.4%; Pred. No. 71;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
| | | | |
Db 524 YEDEVNQ 530

RESULT 13
US-11-134-871-915
; Sequence 915, Application US/11134871
; Publication No. US20060141528A1
; GENERAL INFORMATION:
; APPLICANT: Aebersold, Rudolf H.
; APPLICANT: Zhang, Hui
; TITLE OF INVENTION: Compositions and Methods for
```

```
; TITLE OF INVENTION: Quatification of Serum Glycoproteins
; FILE REFERENCE: 66661-116
; CURRENT APPLICATION NUMBER: US/11/134,871
; CURRENT FILING DATE: 2005-05-20
; PRIOR APPLICATION NUMBER: 60/573,593
; PRIOR FILING DATE: 2004-05-21
; NUMBER OF SEQ ID NOS: 3602
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-134-871-915

Query Match      84.2%; Score 32; DB 7; Length 33;
Best Local Similarity 83.3%; Pred. No. 2.9;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEIN 6
| | | | |
Db 20 YEDEVN 25

RESULT 14
US-10-471-571A-5244
; Sequence 5244, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 5244
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(225)
; OTHER INFORMATION: hypothetical protein
US-10-471-571A-5244

Query Match      84.2%; Score 32; DB 6; Length 225;
Best Local Similarity 71.4%; Pred. No. 23;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
| | | | |
Db 172 YDQINK 178

RESULT 15
US-10-953-349-13561
; Sequence 13561, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 13561
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Glycine max
```



US-10-953-349-13561

Query Match 81.6%; Score 31; DB 6; Length 107;  
Best Local Similarity 71.4%; Pred. No. 16;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7  
|:|||||:  
Db 26 YQDEINE 32

Search completed: July 5, 2006, 20:53:03  
Job time : 4.00971 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 15.767 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-1

Perfect score: 38

Sequence: 1 YEDEINK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5 COMB.pdp.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6 COMB.pdp.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7 COMB.pdp.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H COMB.pdp.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/pCTUS COMB.pdp.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE COMB.pdp.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	7	2	US-09-809-003A-1
2	38	100.0	8	2	US-09-809-003A-12
3	38	100.0	384	2	US-09-949-016-11034
4	38	100.0	443	2	US-09-949-016-10582
5	38	100.0	482	2	US-09-538-092-858
6	38	100.0	483	2	US-09-919-497-79
7	38	100.0	502	2	US-09-949-016-11033
8	38	100.0	551	1	US-09-067-351-2
9	38	100.0	551	2	US-09-360-490-2
10	38	100.0	564	2	US-09-949-016-6628
11	38	100.0	567	2	US-10-104-047-3772
12	38	100.0	569	2	US-09-949-016-11035
13	38	100.0	569	2	US-09-949-016-11036
14	38	100.0	637	2	US-09-949-016-8152
15	38	100.0	643	2	US-09-538-092-844
16	38	100.0	645	2	US-09-919-172-41
17	35	92.1	469	2	US-09-077-606-3
18	35	92.1	476	2	US-09-949-016-9096
19	34	89.5	608	2	US-10-094-749-2446
20	33	86.8	8	2	US-09-809-003A-11
21	33	86.8	216	2	US-09-248-796A-14967
22	33	86.8	546	1	US-09-067-351-1
23	33	86.8	546	2	US-09-360-490-1
24	32	84.2	369	2	US-09-248-796A-18582
25	32	84.2	652	2	US-08-559-896B-2
26	32	84.2	652	2	US-09-351-794A-2

27	32	84.2	1348	2	US-09-949-002-517	Sequence 517, App
28	32	84.2	2196	2	US-10-360-101-259	Sequence 259, App
29	32	84.2	2224	2	US-09-054-272-38	Sequence 38, Appl
30	32	84.2	2224	2	US-09-949-002-292	Sequence 292, App
31	31	81.6	27	2	US-09-270-767-59516	Sequence 59516, A
32	31	81.6	82	2	US-09-270-767-44099	Sequence 44099, A
33	31	81.6	234	2	US-09-270-767-32857	Sequence 32857, A
34	31	81.6	234	2	US-09-270-767-48074	Sequence 48074, A
35	31	81.6	277	2	US-09-538-092-485	Sequence 485, App
36	31	81.6	284	2	US-09-248-796A-17686	Sequence 17686, A
37	31	81.6	286	2	US-09-248-796A-19241	Sequence 19241, A
38	31	81.6	289	2	US-09-248-796A-27946	Sequence 27946, A
39	31	81.6	439	2	US-09-465-558-38	Sequence 38, Appl
40	31	81.6	608	2	US-09-270-767-32937	Sequence 32937, A
41	31	81.6	608	2	US-09-270-767-48154	Sequence 48154, A
42	31	81.6	676	2	US-09-151-189-2	Sequence 2, Appli
43	31	81.6	676	2	US-09-596-794-2	Sequence 2, Appli
44	31	81.6	676	2	US-09-840-762A-2	Sequence 2, Appli
45	31	81.6	676	2	US-10-691-383-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-809-003A-1  
; Sequence 1, Application US/09809003A  
; Patent No. 6673351  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-09-809-003A-1

Query Match 100.0%; Score 38; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7  
Db 1 YEDEINK 7

RESULT 2  
US-09-809-003A-12  
; Sequence 12, Application US/09809003A  
; Patent No. 6673351  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-09-809-003A-12

Query Match 100.0%; Score 38; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 5e+05;

```
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 2 YEDEINK 8

RESULT 3
US-09-949-016-11034
; Sequence 11034, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11034
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11034

Query Match 100.0%; Score 38; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 276 YEDEINK 282

RESULT 4
US-09-949-016-10582
; Sequence 10582, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10582
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10582

Query Match 100.0%; Score 38; DB 2; Length 443;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINK 7
Db 139 YEDEINK 145
```

```
RESULT 5
US-09-538-092-858
; Sequence 858, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 858
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P05787
US-09-538-092-858
```

```
Query Match 100.0%; Score 38; DB 2; Length 482;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 YEDEINK 7
Db 178 YEDEINK 184
```

```
RESULT 6
US-09-919-497-79
; Sequence 79, Application US/09919497
; Patent No. 6773883
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 79
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-79
```

```
Query Match 100.0%; Score 38; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 YEDEINK 7
Db 179 YEDEINK 185
```

```
RESULT 7
US-09-949-016-11033
; Sequence 11033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; PRIOR FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 11033  
;; LENGTH: 502  
;; TYPE: PRT  
;; ORGANISM: Human  
US-09-949-016-11033

Query Match 100.0%; Score 38; DB 2; Length 502;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
DB 198 YEDEINK 204

RESULT 8  
US-09-067-351-2  
; Sequence 2, Application US/09067351  
; Patent No. 5934081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/067,351  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-067-351-2

Query Match 100.0%; Score 38; DB 1; Length 551;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
DB 239 YEDEINK 245

RESULT 9  
US-09-360-490-2  
; Sequence 2, Application US/09360490  
; Patent No. 6221843  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/360,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/067,351  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-360-490-2

Query Match 100.0%; Score 38; DB 2; Length 551;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINK 7  
DB 239 YEDEINK 245

RESULT 10  
US-09-949-016-6628  
; Sequence 6628, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6628
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6628

Query Match          100.0%; Score 38; DB 2; Length 564;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINK 7
Db      253 YEDEINK 259

RESULT 11
US-10-104-047-3772
; Sequence 3772, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. 6943241el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3772
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3772

Query Match          100.0%; Score 38; DB 2; Length 567;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINK 7
Db      235 YEDEINK 241

RESULT 12
US-09-949-016-11035
; Sequence 11035, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11035
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11035

Query Match          100.0%; Score 38; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINK 7
Db      259 YEDEINK 265

RESULT 13
US-09-949-016-11036
; Sequence 11036, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11036
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11036

Query Match          100.0%; Score 38; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINK 7
Db      259 YEDEINK 265

RESULT 14
US-09-949-016-8152
; Sequence 8152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8152
; LENGTH: 637
; TYPE: PRT
```

```
; ORGANISM: Human
US-09-949-016-8152

Query Match      100.0%; Score 38; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 YEDEINK 7
DB      270 YEDEINK 276

RESULT 15
US-09-538-092-844
; Sequence 844, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 844
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number P04264
US-09-538-092-844

Query Match      100.0%; Score 38; DB 2; Length 643;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 YEDEINK 7
DB      269 YEDEINK 275
```

Search completed: July 5, 2006, 19:24:27  
Job time : 16.767 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:20:41 ; Search time 8.25 Seconds  
(without alignments)  
104.964 Million cell updates/sec

Title: US-10-774-928A-2

Perfect score: 45

Sequence: 1 AQVEDIAQK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80.\*  
1: Pirl.\*  
2: Pirl2.\*  
3: Pirl3.\*  
4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	93.3	629	2 A29666	Keratin, 65K type
2	42	93.3	629	2 S42629	Keratin K3 - rabbit
3	41	91.1	524	2 A23518	Keratin, 57K type
4	40	88.9	513	2 S08381	Keratin, 58K type
5	39	86.7	461	2 A43782	Keratin, type II -
6	39	86.7	502	2 A23547	Keratin, type II c
7	38	84.4	384	2 I61769	Keratin 6d, type I
8	38	84.4	534	2 I37942	Keratin 4, type II
9	38	84.4	564	1 KRHUEA	Keratin 6a, type I
10	38	84.4	564	1 KRHUEB	Keratin 6b, type I
11	38	84.4	564	2 I61770	Keratin 6e, type I
12	38	84.4	564	2 I61768	Keratin 6c, type I
13	38	84.4	564	2 I61771	Keratin 6f, type I
14	38	84.4	645	2 A44861	Keratin, 67K type
15	37	82.2	330	2 A60093	Cytokeratin, type
16	37	82.2	483	2 A34720	Keratin 8, type II
17	37	82.2	487	2 JT0407	Keratin 8, type II
18	37	82.2	489	2 S05474	Keratin 8, type II
19	37	82.2	490	2 JS0558	Cytokeratin EndoA
20	37	82.2	520	2 JS0291	Intermediate filam
21	36	80.0	159	2 H83845	stage V sporulatio
22	36	80.0	643	1 KRHU2	Keratin 1, type II
23	35	77.8	174	2 S39790	male-enhanced anti
24	35	77.8	174	2 B34421	male-enhanced anti
25	35	77.8	185	2 A34421	male-enhanced anti
26	35	77.8	368	2 A28825	Keratin, type I no
27	35	77.8	875	2 S70115	ZIP1 protein - yea
28	34	75.6	395	2 D97193	PLP-dependent amin
29	34	75.6	424	2 S14887	peripherin (clone

RESULT 1

A29666

keratin, 65K type II cytoskeletal - human

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 20-Apr-2001

C:Accession: A29666

R:Klinge, E.M.; Sylvestre, Y.R.; Freedberg, I.M.; Blumenberg, M.

J. Mol. Evol. 24, 319-329, 1987

A:Title: Evolution of keratin genes: different protein domains evolve by different pathwa

A:Reference number: A29666, MUID:87254239; PMID:2439698

A:Accession: A29666

A:Molecule type: DNA

A:Residues: 1-629 <KLI>

A:Cross-references: UNIPARC:UPI0000161C35; GB:X05418; NID:934040; PIDN:CAA28991.1; PID:g

92; PIDN:CAA28994.1; PID:g1335193; PIDN:CAA28995.1; PID:g1335194; GB:X05421; NID:g34047;

A>Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 2

A>Note: the complete translation is not annotated in GenBank entries HSKER65A, HSKER65B,

s mistranslated as a Met initiator codon

C:Genetics:

A:Introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 93.3%; Score 42; DB 2; Length 629;

Best Local Similarity 88.9%; Pred. No. 0.96;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQVEDIAQK 9

Db 376 AQVEDIAQK 384

RESULT 2

S42629

keratin K3 - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004

C:Accession: S42629

R:Wu, R.L.; Galvin, S.; Wu, S.K.; Xu, C.; Blumenberg, M.; Sun, T.T.

J. Cell Sci. 105, 303-316, 1993

A:Title: A 300 bp 5'-upstream sequence of a differentiation-dependent rabbit K3 keratin s

A:Reference number: S42629; MUID:94013038; PMID:7691837

A:Accession: S42629

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-629 <WUR>

A:Cross-references: UNIPARC:UPI0000087A39; EMBL:X74371; NID:g433957; PID:

C:Genetics:

A:Introns: 200/3; 274/2; 294/3; 326/3; 381/3; 423/3; 497/2; 509/1

C:Superfamily: cytoskeletal keratin

Query Match 93.3%; Score 42; DB 2; Length 629;

Best Local Similarity 88.9%; Pred. No. 0.96;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQVEDIAQK 9  
| | | | | | | |  
Db 361 AQVEDIAQR 369

RESULT 3  
A23518  
keratin, 57K type II cytoskeletal - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 09-Jul-2004  
C:Accession: A23518  
R:Knapp, B.; Rentrop, M.; Schweizer, J.; Winter, H.  
Nucleic Acids Res. 14, 751-763, 1986  
A:Title: Nonepidermal members of the keratin multigene family: cDNA sequences and in situ  
A:Reference number: A93640; MUID:86120369; PMID:2418416  
A:Accession: A23518  
A:Molecule type: mRNA  
A:Residues: 1-524 <KNA>  
A:Cross-references: UNIPROT:P07744; UNIPARC:UPI000000298E2; GB:X03491; NID:G52784; PIDN:G  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 91.1%; Score 41; DB 2; Length 524;  
Best Local Similarity 88.9%; Pred. No. 1.3;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQVEDIAQK 9  
| | | | | | | |  
Db 322 AQVEDIAK 330

RESULT 4  
S08381  
keratin, 58K type II, cytoskeletal (clone pXenCK55(5/6)) - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 09-Jul-2004  
C:Accession: S08381  
R:Pouquet, B.; Herrmann, H.; Franz, J.K.; Franke, W.W.  
Development 104, 533-548, 1988  
A:Title: Expression of intermediate filament proteins during development of Xenopus laevis  
A:Reference number: S08381; MUID:90032362; PMID:2478354  
A:Accession: S08381  
A:Molecule type: mRNA  
A:Residues: 1-513 <FOU>  
A:Cross-references: UNIPROT:P16878; UNIPARC:UPI0000017151D; EMBL:X14427; NID:G64625; PIDN:G  
C:Superfamily: cytoskeletal keratin

Query Match 88.9%; Score 40; DB 2; Length 513;  
Best Local Similarity 88.9%; Pred. No. 2;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQVEDIAQK 9  
| | | | | | | |  
Db 303 AQVEDIAK 311

RESULT 5  
A43782  
keratin, type II - eastern newt (fragment)  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C:Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A43782  
R:Ferretti, P.; Brookes, J.P.; Brown, R.  
Development 111, 497-507, 1991  
A:Title: A new type II keratin restricted to normal and regenerating limbs and tails is  
A:Reference number: A43782; MUID:91372147; PMID:1716554  
A:Accession: A43782  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-461 <FER>

A;Cross-references: UNIPROT:O42435; UNIPARC:UPI00001774C4; GB:X57671  
C;Superfamily: cytoskeletal keratin

Query Match 86.7%; Score 39; DB 2; Length 461;  
Best Local Similarity 77.8%; Pred. No. 2.9;  
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 AQYEDIAQK 9  
| | | | | | | |  
Db 220 AQYEDVANK 228

RESULT 6  
A23547  
keratin, type II cytoskeletal - African clawed frog  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 09-Jul-2004  
C;Accession: A23547  
R;Franz, J.K.; Franke, W.W.  
Proc. Natl. Acad. Sci. U.S.A. 83, 6475-6479, 1986  
A;Title: Cloning of cDNA and amino acid sequence of a cyto keratin expressed in oocytes of  
A;Reference number: A23547; MUID:86313601; PMID:2428034  
A;Accession: A23547  
A;Molecule type: mRNA  
A;Residues: 1-502 <FRA>  
A;Cross-references: UNIPROT:P08776; UNIPARC:UPI000012DB12; GB:M13811; NID:g214555; PIDN:?  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 86.7%; Score 39; DB 2; Length 502;  
Best Local Similarity 77.8%; Pred. No. 3.2;  
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 AQYEDIAQK 9  
| | | | | | | |  
Db 273 AQYEDVANK 281

RESULT 7  
I61769  
keratin 6d, type II - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
C;Accession: I61769  
R;Takahashi, K.; Pauladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A;Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A;Reference number: A57398; MUID:95355491; PMID:7543104  
A;Accession: I61769  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-384 <RES>  
A;Cross-references: UNIPROT:P48667; UNIPARC:UPI000012DB17; GB:L42602; NID:g914823; GB:L42602; GB:L42608; NID:g914829; GB:L42609; NID:g914830; GB:L42610; NID:g1488252; PIDN:AA86065  
C;Genetics:

A;Gene: KRT6D  
A;Introns: 72/2; 92/3; 124/3; 179/3; 221/3; 295/2; 307/1  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 84.4%; Score 38; DB 2; Length 384;  
Best Local Similarity 77.8%; Pred. No. 3.9;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDIAQK 9  
| | | | | | | |  
Db 159 AQYEEIAQR 167

RESULT 8  
I37942  
keratin 4, type II, cytoskeletal - human  
N;Alternate names: basic cyto keratin; cyto keratin 4

Best Local Similarity 88.9%; Pred. No. 0.96;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQVEDIAQK 9  
| | | | | | | |  
Db 361 AQVEDIAQR 369

RESULT 3  
A23518  
keratin, 57K type II cytoskeletal - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 09-Jul-2004  
C:Accession: A23518  
R:Knapp, B.; Rentrop, M.; Schweizer, J.; Winter, H.  
Nucleic Acids Res. 14, 751-763, 1986  
A:Title: Nonepidermal members of the keratin multigene family: cDNA sequences and in situ  
A:Reference number: A93640; MUID:86120369; PMID:2418416  
A:Accession: A23518  
A:Molecule type: mRNA  
A:Residues: 1-524 <KNA>  
A:Cross-references: UNIPROT:P07744; UNIPARC:UPI000000298E2; GB:X03491; NID:G52784; PIDN:G  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 91.1%; Score 41; DB 2; Length 524;  
Best Local Similarity 88.9%; Pred. No. 1.3;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQVEDIAQK 9  
| | | | | | | |  
Db 322 AQVEDIAK 330

RESULT 4  
S08381  
keratin, 58K type II, cytoskeletal (clone pXenCK55(5/6)) - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 09-Jul-2004  
C:Accession: S08381  
R:Pouquet, B.; Herrmann, H.; Franz, J.K.; Franke, W.W.  
Development 104, 533-548, 1988  
A:Title: Expression of intermediate filament proteins during development of Xenopus laevis  
A:Reference number: S08381; MUID:90032362; PMID:2478354  
A:Accession: S08381  
A:Molecule type: mRNA  
A:Residues: 1-513 <FOU>  
A:Cross-references: UNIPROT:P16878; UNIPARC:UPI0000017151D; EMBL:X14427; NID:G64625; PIDN:G  
C:Superfamily: cytoskeletal keratin

Query Match 88.9%; Score 40; DB 2; Length 513;  
Best Local Similarity 88.9%; Pred. No. 2;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQVEDIAQK 9  
| | | | | | | |  
Db 303 AQVEDIAK 311

RESULT 5  
A43782  
keratin, type II - eastern newt (fragment)  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C:Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A43782  
R:Ferretti, P.; Brookes, J.P.; Brown, R.  
Development 111, 497-507, 1991  
A:Title: A new type II keratin restricted to normal and regenerating limbs and tails is  
A:Reference number: A43782; MUID:91372147; PMID:1716554  
A:Accession: A43782  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-461 <FER>

C;Species: Homo sapiens (man)  
 C;Date: 12-Aug-1996 #sequence\_revision 17-Mar-2000 #text\_change 09-Jul-2004  
 C;Accession: J37942; S01068; S21884; A30186; S31662  
 R;Wanner, R.; Forster, H.H.; Tilmans, I.; Mischke, D.  
 J. Invest. Dermatol. 100, 735-741, 1993  
 A;Title: Allelic variations of human keratins K4 and K5 provide polymorphic markers with  
 A;Reference number: J37942; MUID:93267125; PMID:7684424  
 A;Accession: J37942  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-127, 'I', 129-144 <WAN1>  
 A;Cross-references: UNIPROT:P19013; UNIPARC:UPI000016AB8C; EMBL:X67683; NID:G34020; PIDN:  
 R;Leube, R.E.; Bader, B.L.; Bosch, F.X.; Zimbelmann, R.; Achtsaetter, T.; Franke, W.W.  
 J. Cell Biol. 106, 1249-1261, 1988  
 A;Title: Molecular characterization and expression of the stratification-related cytokin  
 A;Reference number: S01068; MUID:88198369; PMID:2452170  
 A;Accession: S01068  
 A;Molecule type: mRNA  
 A;Residues: 'S', 128-534 <LEU>  
 A;Cross-references: UNIPARC:UPI000016ABA3; EMBL:X07695; NID:G34072; PIDN:CAA30534.1; PID  
 R;Wanner, R.; Tilmans, I.; Mischke, D.  
 submitted to the EMBL Data Library, July 1991  
 A;Reference number: S21884  
 A;Accession: S21884  
 A;Molecule type: DNA  
 A;Residues: 411-534 <WAN2>  
 A;Cross-references: UNIPARC:UPI000016ABA5; EMBL:X61028; NID:G34076; PIDN:CAA43362.1; PID  
 A;Experimental source: allele K4a  
 C;Genetics:  
 A;Gene: GDB:KRT4; CYK4  
 A;Cross-references: GDB:120697; OMIM:123940  
 A;Map position: 12p11.2-12q11  
 A;Introns: 463/2; 475/1  
 C;Superfamily: cytoskeletal keratin  
 C;Keywords: coiled coil; intermediate filament  
 F;1-152/Domain: head #status predicted <HEA>  
 F;153-464/Domain: helical rod #status predicted <ROD>  
 F;465-534/Domain: tail #status predicted <TAI>

Query Match 84.4%; Score 38; DB 2; Length 534;  
 Best Local Similarity 77.8%; Pred. No. 5.6;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9  
 |||||:  
 Db 327 AQVEEIAQR 335

RESULT 9  
 KRHUEA  
 keratin 6a, type II - human  
 N;Alternate names: 56-kDa type II keratin; keratin cytoskeletal  
 C;Species: Homo sapiens (man)  
 C;Date: 15-Nov-1984 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: A57398; A02944  
 R;Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
 J. Biol. Chem. 270, 18581-18592, 1995  
 A;Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
 A;Reference number: A57398; MUID:95355491; PMID:7543104  
 A;Accession: A57398  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-564 <TAK>  
 A;Cross-references: UNIPROT:P02538; UNIPARC:UPI000013CD4C; GB:L42575; NID:G908769; GB:L4  
 74; GB:L42581; NID:G908775; GB:L42582; NID:G908776; GB:L42583; NID:G908777; PIDN:AAC4176  
 R;Hanukoglu, I.; Fuchs, E.  
 Cell 33, 915-924, 1983  
 A;Title: The cDNA sequence of a type II cytoskeletal keratin reveals constant and variab  
 A;Reference number: A02944; MUID:83259278; PMID:6191871  
 A;Accession: A02944  
 A;Molecule type: mRNA  
 A;Residues: 208-394, 'S', 396-564 <HAN>  
 A;Cross-references: UNIPARC:UPI000016AB87; GB:J00269; NID:G34068; PIDN:CAA24760.1; PID:9

C;Comment: The cytoskeletal and microfibrillar keratins are classified into two types, ty  
 atin IF protein subunit appears to be a heterotetramer of two type I and two type II prot  
 C;Genetics:  
 A;Gene: GDB:KRT6A  
 A;Cross-references: GDB:128111; OMIM:148041  
 A;Map position: 12q12-12q21  
 A;Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1  
 C;Superfamily: cytoskeletal keratin  
 C;Keywords: coiled coil; intermediate filament  
 F;2-163/Domain: head <HED>  
 F;2-163/Region: E1 and V1 subdomains  
 F;127-163/Region: H1 subdomain  
 F;162-476/Domain: rod <ROD>  
 F;199-210/Region: linker 1  
 F;211-311/Region: coil 1B  
 F;312-328/Region: linker 12  
 F;329-347/Region: coil 2A  
 F;348-355/Region: linker 2  
 F;356-476/Region: coil 2B  
 F;414/Region: stutter  
 F;477-564/Domain: tail <END>  
 F;477-496/Region: H2 subdomain  
 F;497-564/Region: V2 and E2 subdomains

Query Match 84.4%; Score 38; DB 1; Length 564;  
 Best Local Similarity 77.8%; Pred. No. 6;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9  
 |||||:  
 Db 339 AQVEEIAQR 347

RESULT 10  
 KRHUEB  
 keratin 6b, type II - human  
 N;Alternate names: 56K type II keratin; keratin, cytoskeletal  
 C;Species: Homo sapiens (man)  
 C;Date: 04-Dec-1986 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: I61767; A02945  
 R;Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
 J. Biol. Chem. 270, 18581-18592, 1995  
 A;Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
 A;Reference number: A57398; MUID:95355491; PMID:7543104  
 A;Accession: I61767  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-564 <RES>  
 A;Cross-references: UNIPROT:P04259; UNIPARC:UPI0000167B38; GB:L42584; NID:G908780; GB:L42  
 85; GB:L42590; NID:G908786; GB:L42591; NID:G908787; GB:L42592; NID:G908788; PIDN:AAC4176  
 R;Tyner, A.L.; Eichman, M.J.; Fuchs, E.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4683-4687, 1985  
 A;Title: The sequence of a type II keratin gene expressed in human skin: conservation of  
 A;Reference number: A02945; MUID:85270392; PMID:2410904  
 A;Accession: A02945  
 A;Molecule type: DNA  
 A;Residues: 2-88, 'AG', 91-115, 'PA', 118, 'LC', 122-158, 'IG', 161-254, 'V', 256-564 <TYN>  
 A;Cross-references: UNIPARC:UPI0000173D58; GB:M11229; GB:L00205; NID:G186714; PIDN:AAAS94  
 A;Note: Initiator Met not shown  
 C;Comment: There are two types of cytoskeletal and microfibrillar keratin, I (acidic) and  
 bligate heteropolymers, composed of complexes formed by the aggregation of at least one  
 C;Genetics:  
 A;Gene: GDB:KRT6B  
 A;Cross-references: GDB:128113; OMIM:148042  
 A;Map position: 12pter-12qter  
 A;Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1  
 C;Superfamily: cytoskeletal keratin  
 C;Keywords: coiled coil; intermediate filament  
 F;2-163/Domain: head <HED>  
 F;2-163/Region: E1 and V1 subdomains  
 F;128-163/Region: H1 subdomain  
 F;164-476/Domain: rod <ROD>  
 F;164-198/Region: coil 1A

F:199-210/Region: linker 1  
 F:211-311/Region: coil 1B  
 F:312-328/Region: linker 12  
 F:329-347/Region: coil 2A  
 F:348-355/Region: linker 2  
 F:356-476/Region: coil 2B  
 F:414/Region: stutter  
 F:477-564/Domain: tail <END>  
 F:477-496/Region: H2 subdomain  
 F:497-564/Region: V2 and E2 subdomains

Query Match 84.4%; Score 38; DB 1; Length 564;

Best Local Similarity 77.8%; Pred. No. 6;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9

|||||:

Db 339 AQYEBIAQR 347

RESULT 11

I61770

Keratin 6e, type II - human

C:Species: Homo sapiens (man)

C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C:Accession: I61770

R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.

J. Biol. Chem. 270, 18581-18592, 1995

A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A:Reference number: A57398; MUID:95355491; PMID:7543104

A:Accession: I61770

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-564 <RES>

A:Cross-references: UNIPROT:P48668; UNIPARC:UPI0000167E31; GB:L42611; NID:G908802; PIDN:

C:Genetics:

A:Gene: KRT6E

A:Note: this may not be a distinct gene

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 84.4%; Score 38; DB 2; Length 564;

Best Local Similarity 77.8%; Pred. No. 6;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9

|||||:

Db 339 AQYEBIAQR 347

RESULT 12

I61768

Keratin 6c, type II - human

C:Species: Homo sapiens (man)

C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C:Accession: I61768

R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.

J. Biol. Chem. 270, 18581-18592, 1995

A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A:Reference number: A57398; MUID:95355491; PMID:7543104

A:Accession: I61768

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-564 <RES>

A:Cross-references: UNIPROT:P48666; UNIPARC:UPI0000167E0; GB:L42593; NID:G908791; GB:L4

96; GB:L42599; NID:G908797; GB:L42600; NID:G908798; GB:L42601; NID:G908799; PIDN:AAC4176

C:Genetics:

A:Gene: KRT6C

A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 84.4%; Score 38; DB 2; Length 564;

Best Local Similarity 77.8%; Pred. No. 6;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9

|||||:

Db 339 AQYEBIAQR 347

RESULT 13

I61771

Keratin 6f, type II - human

C:Species: Homo sapiens (man)

C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C:Accession: I61771

R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.

J. Biol. Chem. 270, 18581-18592, 1995

A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A:Reference number: A57398; MUID:95355491; PMID:7543104

A:Accession: I61771

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-564 <RES>

A:Cross-references: UNIPROT:P48669; UNIPARC:UPI000013CD50; GB:L42612; NID:G908804; PIDN:

C:Genetics:

A:Gene: KRT6F

A:Note: this may not be a distinct gene

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 84.4%; Score 38; DB 2; Length 564;

Best Local Similarity 77.8%; Pred. No. 6;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9

|||||:

Db 339 AQYEBIAQR 347

RESULT 14

A44861

Keratin, 67K type II epidermal - human

C:Alternate names: cytokeratin 2, CK 2; epidermal cytokeratin 2

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004

C:Accession: A44861

R:Collin, C.; Moll, R.; Kubicka, S.; Ouhayoun, J.P.; Franke, W.W.

Exp. Cell Res. 202, 132-141, 1992

A:Title: Characterization of human cytokeratin 2, an epidermal cytoskeletal protein synt

A:Reference number: A44861; MUID:92380238; PMID:1380918

A:Accession: A44861

A:Molecule type: mRNA

A:Residues: 1-645 <COL>

A:Cross-references: UNIPROT:P35908; UNIPARC:UPI000012DB03; GB:S43646

A:Experimental source: epidermis

A:Note: the authors translated the codon GGC for residue 146 as Cys

A:Note: sequence extracted from NCBI backbone (NCBIN:112351)

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

Query Match 84.4%; Score 38; DB 2; Length 645;

Best Local Similarity 77.8%; Pred. No. 6.9;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9

|||||:

Db 360 AQYEBIAQR 368

RESULT 15

A60093

Cytokeratin, type II, early ectodermal - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 30-Sep-1993

C;Accession: A60093  
R;Charlebois, T.S.; Spencer, D.H.; Tarkington, S.K.; Henry, J.J.; Grainger, R.M.  
Development 108, 33-45, 1990  
A;Title: Isolation of a chick cytokeatin cDNA clone indicative of regional specialization  
A;Reference number: A60093; MUID:90276238; PMID:1693557  
A;Accession: A60093  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-390 <CHA>  
A;Cross-references: UNIPARC:UPI00001774C6  
C;Superfamily: cytoskeletal keratin  
C;Keywords: intermediate filament

Query Match 82.2%; Score 37; DB 2; Length 390;  
Best Local Similarity 77.8%; Pred. No. 6.5;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 187 AQYEDIANR 195

Search completed: July 5, 2006, 19:25:11  
Job time : 8.25 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:01:48 ; Search time 104.625 Seconds  
(without alignments)  
79.571 Million cell updates/sec

Title: US-10-774-928A-2

Perfect score: 45

Sequence: 1 AQVEDIAQK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	594	Q3UV17_MOUSE	Q3UV17 mus musculus
2	42	93.3	278	Q9CV72_MOUSE	Q9CV72 mus musculus
3	42	93.3	552	K2C6A_MOUSE	F50446 mus musculus
4	42	93.3	554	Q3UV11_MOUSE	Q3UV11 mus musculus
5	42	93.3	561	K2C6B_MOUSE	Q92331 mus musculus
6	42	93.3	581	Q6IFZ5_RAT	O6IFZ5 rattus norv
7	42	93.3	629	K2C3_HUMAN	P12035 homo sapien
8	42	93.3	629	K2C3_RABIT	Q29426 oryctolagus
9	41	91.1	525	K2C4_MOUSE	F07744 mus musculus
10	41	91.1	580	Q98UJ3_XENLA	Q98UJ3 xenopus lae
11	41	91.1	619	Q6EIV9_CANFA	O6EIV9 canis famil
12	41	91.1	633	Q6EIZ1_CANFA	O6EIZ1 canis famil
13	40	88.9	496	Q24234_NOTVI	O42434 notophthalm
14	40	88.9	508	Q6NVR6_XENTR	Q6NVR6 xenopus tro
15	40	88.9	511	Q90ZF7_RANCA	Q90ZF7 rana catesb
16	40	88.9	512	K2C5_XENLA	P16878 xenopus lae
17	40	88.9	537	Q91219_ONCMY	Q91219 oncorhynch
18	39	86.7	459	Q7SYF7_ACIBE	Q7SYF7 acipenser b
19	39	86.7	461	Q24235_NOTVI	O42435 notophthalm
20	39	86.7	473	Q4RMX0_TETNG	Q4RMX0 tetraodon n
21	39	86.7	487	Q7SYF6_ACIBE	Q7SYF6 acipenser b
22	39	86.7	502	K2C8_XENLA	P08776 xenopus lae
23	39	86.7	507	Q7ZXU4_XENLA	Q7ZXU4 xenopus lae
24	39	86.7	538	Q90VH3_ONCMY	Q90VH3 oncorhynch
25	38	84.4	384	K2C6D_HUMAN	P48667 homo sapien
26	38	84.4	497	Q5K2N3_PROAT	O5K2N3 protopterus
27	38	84.4	534	K2C4_HUMAN	P19013 homo sapien
28	38	84.4	536	Q6PVZ4_CHICK	Q6PVZ4 gallus gall
29	38	84.4	539	Q6NXH9_MOUSE	Q6NXH9 mus musculus
30	38	84.4	553	Q6IG03_RAT	O6IG03 rattus norv
31	38	84.4	563	K2C6A_HUMAN	P02538 homo sapien

32	38	84.4	563	1	K2C6B_HUMAN	P04259 homo sapien
33	38	84.4	563	1	K2C6C_HUMAN	P48666 homo sapien
34	38	84.4	563	1	K2C6E_HUMAN	P48668 homo sapien
35	38	84.4	564	2	Q2TAZ9_HUMAN	Q2TAZ9 homo sapien
36	38	84.4	639	2	Q4VAQ2_HUMAN	Q4VAQ2 homo sapien
37	38	84.4	645	1	K22E_HUMAN	P35908 homo sapien
38	38	84.4	653	2	Q9SLG9_CANFA	Q9SLG9 canis famil
39	37	82.2	79	2	Q63280_RAT	Q63280 rattus norv
40	37	82.2	176	2	Q4JU123_ICTPU	Q4JU123 ictalurus p
41	37	82.2	279	2	Q7L4M3_HUMAN	Q7L4M3 homo sapien
42	37	82.2	305	2	Q4SY71_TETNG	Q4SY71 tetraodon n
43	37	82.2	346	2	Q4QV72_SPAAU	Q4QV72 sparus aura
44	37	82.2	365	2	Q96910_HUMAN	Q96910 homo sapien
45	37	82.2	378	1	CAAL_BDEBA	Q6mkw4 bdellovibri

## ALIGNMENTS

### RESULT 1

Q3UV17_MOUSE	PRELIMINARY; PRT; 594 AA.
AC Q3UV17;	
DT 11-OCT-2005,	integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006,	entry version 6.
DE Adult female vagina cDNA, RIKEN full-length enriched library,	
GN Name=2310001L23Rik;	
OS Mus musculus (Mouse).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;	
OC Muridae; Murinae; Mus.	
OX NCBI_TaxID=10090;	
RN [1]	
RP NUCLEOTIDE SEQUENCE.	
RC STRAIN=C57BL/6J; TISSUE=Vagina;	
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;	
RA Carninci P., Hayashizaki Y.;	
RT "High-efficiency full-length cDNA cloning.";	
RL Methods Enzymol. 303:19-44(1999).	
RN [2]	
RP NUCLEOTIDE SEQUENCE.	
RC STRAIN=C57BL/6J; TISSUE=Vagina;	
RX PubMed=16141072; DOI=10.1126/science.1112014;	
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,	
Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,	
Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,	
Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,	
Ambei-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,	
Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,	
Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,	
Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,	
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,	
Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,	
Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,	
Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,	
Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,	
Jakt M., Kanincki A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,	
Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,	
Kurochkin I.V., Lareau L.P., Lazarevic D., Lipovich L., Liu J.,	
Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,	
Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,	
Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,	
Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,	
Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,	
Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,	
Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Sheng Y.,	
Schombach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,	
Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,	
Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,	
Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,	
Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,	

RA Yamamichi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlesed C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563 (2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX PubMed16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566 (2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia K., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanae A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konggaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Nunata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shinada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlesed C., Wang Y., Watanabe Y., Wells C.,  
 RA Wu Ming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs";  
 RL Nature 420:563-573 (2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischnmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,

RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690 (2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630 (2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771 (2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
 RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AK137676; BAE23456.1; -; mRNA.  
 DR MGI; MGI:1924305; 2310001L23Rik.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR003054; Keratin\_II.  
 DR Pfam; PF00038; Filament; I.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Coiled coil; Intermediate filament; Keratin.  
 SQ SEQUENCE 594 AA; 62845 MW; 0099982B496DF612 CRC64;  
 Query Match 100.0%; Score 45; DB 2; Length 594;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AQYEDIAQK 9  
 DB 343 AQYEDIAQK 351  
 RESULT 2  
 Q9CV72\_MOUSE  
 ID Q9CV72\_MOUSE PRELIMINARY; PRT; 278 AA.  
 AC Q9CV72;  
 DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 22.  
 DE Adult male tongue cDNA, RIKEN full-length enriched library,  
 DE clone:2310008N03 product:keratin complex 2, basic, gene 6b, full  
 DE insert sequence. (fragment).  
 GN Name=Krt2-6b;



OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
ON [1]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Tongue;  
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RX Carninci P., Hayashizaki Y.;  
RA "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Tongue;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impombato A., Apweiler R., Attalaya R.N., Bailey T.L., Banerji M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R., Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G., Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M., Georgii-Hemmen P., Gieger T., Gojorori T., Green R.E., Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V., Lareau L.P., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y., Orlandini V., Pang K.C., Pavan W.J., Pavoni G., Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M., Rost B., Ruan Y., Salzberg S.L., Sadelain A., Schneider C., Schonbach C., Sekiguchi K., Sempere C.A., Seno S., Sessa L., Sheng Y., Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B., Sperling S., Stupka E., Sugtara K., Sultana R., Takenawa Y., Taki K., Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K., Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C., Grimmond S.M., Tesdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Lida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashina T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N., Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S., Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Tongue;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Tongue;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojorori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Ciothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kasawawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltakis L., Marchionni L., Mckenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reid J., Ring B.Z., Ringwald M., Sadelain A., Schneider C., Sempere C.A., Setou M., Shimada K., Sultana R., Takenawa Y., Taylor M.S., Tesdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Tongue;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Iwama M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojorori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fetschmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bernaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schonbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Tongue;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Tongue;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsumi T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Iwama K., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Tongue;

```

RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayase N., Hiramoto K., Hiraoka T., Horii F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuayama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saïto H., Saïto R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi P., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
DR EMBL; AK009241; BAB26162.1; -; mRNA.
DR HSSP; P08670; 1GK4.
DR Ensembl; ENSMUSG00000023041; Mus musculus.
DR MGI; MGI:1333768; Krt2-6b.
DR GO; GO:0046021; C:integral to membrane; RCA.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR GO; GO:0045109; P:intermediate filament organization; IGI.
DR GO; GO:0031424; P:keratinization; IGI.
DR GO; GO:0002009; P:morphogenesis of an epithelium; IGI.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR18893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
FT NON_TER
FT SEQUENCE 278 AA; 30011 MW; BC641612D2959191 CRC64;
Query Match 93.3%; Score 42; DB 2; Length 278;
Best Local Similarity 88.9%; Pred. No. 4, 5;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQYEDIAQK 9
Db |||||||:
52 AQYEDIAQR 60
RESULT 3
K2C6A_MOUSE
ID K2C6A_MOUSE STANDARD; PRT; 552 AA.
AC P50446; Q9Z332;
DT 01-OCT-1996, integrated into UniProtKB/Swiss-Prot.
DT 01-MAR-2005, sequence version 2.
DT 21-FEB-2006, entry version 42.
DE Keratin, type II cytoskeletal 6A (Cytokeratin-6A) (CK 6A) (K6a
GN Name=Krt6a; Synonyms=Ker2, Krt2-6, Krt2-6a, Krt6;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC TISSUE=Epidermis;
RX MEDLINE=85014838; PubMed=6207530;
RA Steinert P.M., Parry D.A.D., Racoosin E.L., Idler W.W., Steven A.C.,
RA Trus B.L., Roop D.R.;
RT "The complete cDNA and deduced amino acid sequence of a type II mouse
RT epidermal keratin of 60,000 Da: analysis of sequence differences
RT between type I and type II keratins.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5709-5713 (1984).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], TISSUE SPECIFICITY, AND INDUCTION.
RC STRAIN=129/Sv; TISSUE=Skin;

```

```

RX Takahashi K., Yan B., Yamanishi K., Imamura S., Coulombe P.A.;
RT "The two functional keratin 6 genes of mouse are differentially
RL regulated and evolved independently from their human orthologs.";
RN Genomics 53:170-183 (1998).
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Jaw, and Limb;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RP NUCLEOTIDE SEQUENCE OF 527-552.
RX PubMed=1713533;
RA Finch J., Andrews K., Krieg P., Furstenberger G., Slaga T.,
RA Ootsuyama A., Tanooka H., Bowden G.T.;
RT "Identification of a cloned sequence activated during multi-stage
RT carcinogenesis in mouse skin.";
RL Carcinogenesis 12:1519-1522 (1991).
CC -!- FUNCTION: There are two types of cytoskeletal and microfibrillar
CC keratin: I (acidic; 40-55 kDa) [K9 to K20] and II (neutral to
CC basic; 56-70 kDa) [K1 to K8]. Both a basic and an acidic keratin
CC are required for filament assembly.
CC -!- SUBUNIT: Heterodimer of a type I and a type II keratin. KRT6
CC isomers associate with KRT16 and/or KRT17.
CC -!- TISSUE SPECIFICITY: Predominates in the adult trunk skin, tongue,
CC trachea/esophagus and eye. In adult skin, localization is
CC restricted to hair follicles, where it is localized predominantly
CC in the outer root sheath.
CC -!- INDUCTION: With the exception of specific body sites, keratin 6
CC expression is induced under conditions of epithelial
CC hyperproliferation such as wound healing, certain skin diseases,
CC cancer, and by treatment of the skin with the phorbol ester PMA.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
DR EMBL; K02108; AAA39395.1; -; mRNA.
DR EMBL; AB012033; BAA34178.1; -; Genomic_DNA.
DR EMBL; BC080820; AAB80820.1; -; mRNA.
DR FIR; I59009; I59009.
DR HSSP; P08670; 1GK7.
DR IntAct; P50446; -.
DR Ensembl; ENSMUSG00000058354; Mus musculus.
DR MGI; MGI:1100845; Krt2-6a.
DR GO; GO:0005882; C:intermediate filament; ISS.
DR GO; GO:0005198; P:cytoskeletal molecule activity; ISS.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR InterPro; IPR001664; IF.
DR PANTHER; PTHR18893:SF5; Keratin_II.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.

```

KW Coiled coil; Intermediate filament; Keratin; Phosphorylation.  
 FT INIT\_MET 0 By similarity.  
 FT CHAIN 1 552 Keratin, type II cytoskeletal 6A.  
 FT /FTID=PRO\_0000063736.  
 FT REGION 1 150 Head.  
 FT REGION 151 460 Rod.  
 FT REGION 151 186 Coil 1A.  
 FT REGION 187 205 Linker 1.  
 FT REGION 206 297 Linker 1B.  
 FT REGION 298 321 Linker 12.  
 FT REGION 322 460 Coil 2.  
 FT REGION 461 552 Tail.  
 FT SITE 402 Stutter.  
 FT MOD\_RES 57 57 Phosphoserine (By similarity).  
 FT CONFLICT 23 23 P -> L (in Ref. 1).  
 FT CONFLICT 120 120 P -> L (in Ref. 1).  
 FT CONFLICT 172 172 L -> M (in Ref. 1).  
 FT CONFLICT 179 179 L -> M (in Ref. 1).  
 FT CONFLICT 190 190 G -> D (in Ref. 1).  
 FT CONFLICT 198 198 L -> M (in Ref. 1).  
 FT CONFLICT 223 224 LD -> MN (in Ref. 1).  
 FT CONFLICT 232 239 DTVEDYKS -> ELVEELRN (in Ref. 1).  
 FT CONFLICT 250 250 A -> D (in Ref. 1).  
 FT CONFLICT 311 311 D -> V (in Ref. 1).  
 FT CONFLICT 317 317 D -> V (in Ref. 1).  
 FT CONFLICT 329 331 YED -> FEV (in Ref. 1).  
 FT CONFLICT 343 343 W -> L (in Ref. 1).  
 FT CONFLICT 431 431 R -> M (in Ref. 1).  
 FT CONFLICT 437 437 Q -> H (in Ref. 1).  
 FT CONFLICT 497 497 L -> M (in Ref. 1).  
 FT SEQUENCE 552 AA; 59204 MW; DEB02A645EBA8DAF CRC64;

Query Match 93.3%; Score 42; DB 1; Length 552;  
 Best Local Similarity 88.9%; Pred. No. 9.5;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQVEDIAQK 9  
 DB 327 AQVEDIAQK 335

## RESULT 4

Q3UV11\_MOUSE  
 ID Q3UV11\_MOUSE PRELIMINARY; PRT; 554 AA.  
 AC Q3UV11;  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 11-OCT-2005, sequence version 1.  
 DE 07-FEB-2006, entry version 5.  
 DE Adult female vagina cDNA, RIKEN full-length enriched library,  
 DE clone:930033J10 product:keratin complex 2, basic, gene 6b, full  
 DE insert sequence.  
 GN Name=Krt2-6b;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridea; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX PubMed=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RA "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44 (1999).  
 RN [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aicinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,

RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dall'Ympile B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RA "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563 (2005).  
 RN [3]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566 (2005).  
 RN [4]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Vagina;  
 RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schraml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Naglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wyszaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayashizaki Y.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of

60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Vagina; DOI=10.1038/35055500;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Glessi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guetincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,  
 RA Hayashizaki Y.;  
 RL "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690 (2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630 (2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Vagina; DOI=10.1101/gr.152600;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura K., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771 (2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Vagina;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
 RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 CC EMBL: AK137688; BAE23462.1; -; mRNA.  
 DR MGI: 1333768; Krt2-6b.  
 DR GO: 00016021; C: integral to membrane; RCA.  
 DR GO: 0005739; C: mitochondrion; IDA.  
 DR GO: 00045109; P: intermediate filament organization; IGI.  
 DR GO: 0031424; P: keratinization; IGI.  
 DR GO: 0002009; P: morphogenesis of an epithelium; IGI.  
 DR InterPro: IPR001664; IF.  
 DR InterPro: IPR003054; Keratin\_II.

DR Pfam: PF00038; Filament; 1.  
 DR PRINTS: PR01276; TYPE2KERATIN.  
 DR PROSITE: PS00226; IF; 1.  
 KW Intermediate filament; Keratin.  
 SQ SEQUENCE 554 AA; 59526 MW; 440D8AE398488AA CRC64;  
 Query Match 93.3%; Score 42; DB 2; Length 554;  
 Best Local Similarity 88.9%; Pred. No. 9.5;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AQYEDIAQK 9  
 Db 328 AQYEDIAQK 336  
 RESULT 5  
 K2C6B MOUSE STANDARD; PRT; 561 AA.  
 ID Q92331;  
 AC Q92331;  
 DT 15-FEB-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 15-FEB-2005, sequence version 2.  
 DT 07-FEB-2006, entry version 33.  
 DE Keratin, type II cytoskeletal 6B (Cytokeratin-6B) (CK 6B) (K6b  
 DE keratin) (Keratin-6 beta) (mk6-beta).  
 GN Name=Krt6b; Synonyms=K6-beta, Krt2-6b;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Murioidea; Muridae; Murinae; Mus.  
 RN NCBI\_TaxID=10090;  
 RX STRAIN=129/SV; TISSUE=Skin;  
 RX MEDLINE=95009332; PubMed=9790766; DOI=10.1006/geno.1998.5476;  
 RA Takahashi K., Yan B., Yamanishi K., Imamura S., Coulombe P.A.;  
 RT "The two functional keratin 6 genes of mouse are differentially  
 RT regulated and evolved independently from their human orthologs.";  
 RL Genomics 53:170-183 (1998).  
 CC -!- FUNCTION: There are two types of cytoskeletal and microfibrillar  
 CC keratin: I (acidic; 40-55 kDa) [K9 to K20] and II (neutral to  
 CC basic; 56-70 kDa) [K1 to K8]. Both a basic and an acidic keratin  
 CC are required for filament assembly.  
 CC -!- SUBUNIT: Heterodimer of a type I and a type II keratin. KRT6  
 CC isomers associate with KRT16 and/or KRT17.  
 CC -!- TISSUE SPECIFICITY: Expressed in adult epithelia including the  
 CC tongue, esophagus/trachea, eye and skin. Localized preferentially  
 CC to the suprabasal layers of thickened epidermis in injured and  
 CC chemically treated skin.  
 CC -!- INDUCTION: By injury, and treatment of the skin with the phorbol  
 CC ester PMA.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 CC EMBL: AB012042; BAA34179.1; -; Genomic\_DNA.  
 DR HSSP: P08670; 1GK7.  
 DR Ensembl: ENSMUSG00000023041; Mus musculus.  
 DR MGI: 1333768; Krt2-6b.  
 DR GO: 00016021; C: integral to membrane; TAS.  
 DR GO: 0005882; C: intermediate filament; ISS.  
 DR GO: 0005739; C: mitochondrion; IDA.  
 DR GO: 0005198; P: structural molecule activity; ISS.  
 DR GO: 0007010; P: cytoskeleton organization and biogenesis; ISS.  
 DR InterPro: IPR001664; IF.  
 DR InterPro: IPR003054; Keratin\_II.  
 DR PANTHER: PTHR1893; Sfs; Keratin\_II; 1.  
 DR Pfam: PF00038; Filament; 1.  
 DR PRINTS: PR01276; TYPE2KERATIN.  
 DR PROSITE: PS00226; IF; 1.  
 KW Coiled coil; Intermediate filament; Keratin; Phosphorylation.  
 FT INIT MET 0 0 By similarity.  
 FT CHAIN 1 561 Keratin, type II cytoskeletal 6B.

FT REGION 1 158 /FTID=PRO\_0000063737.  
FT REGION 159 468 Head.  
FT REGION 159 194 Rod.  
FT REGION 159 194 Coil 1A.  
FT REGION 195 213 Linker 1.  
FT REGION 214 305 Coil 1B.  
FT REGION 306 329 Linker 12.  
FT REGION 330 468 Coil 2.  
FT REGION 469 561 Tail.  
FT SITE 410 410 Stutter.  
FT MOD RES 57 57 Phosphoserine (By similarity).  
SQ SEQUENCE 561 AA; 60191 MW; 65B71CAB7F9564E1 CRC64;

Query Match 93.3%; Score 42; DB 1; Length 561;  
Best Local Similarity 88.9%; Pred. No. 9.6;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9  
Db 335 AQVEDIAQR 343  
|||||:

RESULT 6  
Q6IFZ5\_RAT PRELIMINARY; PRT; 581 AA.  
AC Q6IFZ5; integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 13.  
DE Type II keratin Kb9.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN NCBI\_TaxID=10116;  
RP NUCLEOTIDE SEQUENCE  
RX MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;  
RA Hesse M., Zmek A., Weber K., Magin T.M.;  
RT "Comprehensive analysis of keratin gene clusters in humans and rodents.";  
RL Eur. J. Cell Biol. 83:19-26(2004).  
CC -I- MISCELLANEOUS: The sequence shown here is derived from an EMBL/GenBank/DBJ third party annotation (TPA) entry.  
CC -I- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BK003994; DAA02235.1; -; mRNA.  
DR Ensembl; ENSRNOG00000031785; Rattus norvegicus.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR InterPro; IPR003054; Keratin\_II.  
DR PANTHER; PTHR1893.SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 581 AA; 61758 MW; 13B0F3D23BDF4E49 CRC64;

Query Match 93.3%; Score 42; DB 2; Length 581;  
Best Local Similarity 88.9%; Pred. No. 10;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AQVEDIAQK 9  
Db 337 AQVEDIAQR 345  
|||||:

RESULT 7  
K2C3\_HUMAN STANDARD; PRT; 629 AA.  
ID K2C3\_HUMAN  
AC P12035; Q701L8;  
DT 01-OCT-1989, integrated into UniProtKB/Swiss-Prot.  
DT 08-NOV-2005, sequence version 2.  
DT 07-FEB-2006, entry version 48.  
DE Keratin, type II cytoskeletal 3 (Cytokeratin-3) (CK-3) (Keratin-3) (K3) (65 kDa cyokeratin).  
DE Name=KRT3;  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN NCBI\_TaxID=9606;  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RX MEDLINE=87254239; PubMed=2439698;  
RA Klinge E.M., Sylvestre Y.R., Freedberg I.M., Blumenberg M.;  
RT "Evolution of keratin genes: different protein domains evolve by different pathways.";  
RL J. Mol. Evol. 24:319-329(1987).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [mRNA] OF 16-629.  
RC TISSUE=Eye;  
RX PubMed=15737194; DOI=10.1111/j.0022-202X.2004.23530.x;  
RA Rogers M.A., Edler L., Winter H., Langbein L., Beckmann I., Schweizer J.;  
RT "Characterization of new members of the human type II keratin gene family and a general evaluation of the keratin gene domain on chromosome 12q13.13.";  
RL J. Invest. Dermatol. 124:536-544(2005).  
RN [3]  
RP VARIANT MCD LYS-509.  
RX MEDLINE=97315826; PubMed=9171831;  
RA Irvine A.D., Corden L.D., Swensson O., Swensson B., Moore J.E., Frazer D.G., Smith F.J.D., Knowlton R.G., Christophers E., Rochels R., Utto J., McLean W.H.I.;  
RT "Mutations in cornea-specific keratin K3 or K12 genes cause Meesmann's corneal dystrophy.";  
RL Nat. Genet. 16:184-187(1997).  
CC -I- SUBUNIT: Heterotetramer of two type I and two type II keratins.  
CC Keratin-3 associates with keratin-12.  
CC TISSUE SPECIFICITY: Cornea specific.  
CC -I- DISEASE: Defects in KRT3 are a cause of Meesmann corneal dystrophy (MCD) [MIM:122100]. MCD is an autosomal dominant disease that causes fragility of the anterior corneal epithelium. Patients are usually asymptomatic until adulthood when rupture of the corneal microcysts may cause erosions, producing clinical symptoms such as photophobia, contact lens intolerance and intermittent diminution of visual acuity. Rarely, subepithelial scarring causes irregular corneal astigmatism and permanent visual impairment. Histological examination shows a disorganized and thickened epithelium with widespread cytoplasmic vacuolization and numerous small, round, debris-laden intraepithelial cysts.  
CC -I- MISCELLANEOUS: There are two types of cytoskeletal and microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II (neutral to basic; 56-70 kDa) [K1 to K8].  
CC -I- SIMILARITY: Belongs to the intermediate filament family.  
CC -I- DATABASE: NCBIs-Human Intermediate Filament Mutation Database; WWW="http://www.interfil.org".  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; X05418; CAA28991.1; -; Genomic DNA.  
DR EMBL; X05419; CAA28992.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05420; CAA28993.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05421; CAA28994.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05422; CAA28995.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05423; CAA28996.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05424; CAA28997.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05425; CAA28998.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05426; CAA28999.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05427; CAA29000.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05428; CAA29001.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05429; CAA29002.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05430; CAA29003.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05431; CAA29004.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05432; CAA29005.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05433; CAA29006.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05434; CAA29007.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05435; CAA29008.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05436; CAA29009.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05437; CAA29010.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05438; CAA29011.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05439; CAA29012.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05440; CAA29013.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05441; CAA29014.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05442; CAA29015.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05443; CAA29016.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05444; CAA29017.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05445; CAA29018.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05446; CAA29019.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05447; CAA29020.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05448; CAA29021.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05449; CAA29022.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05450; CAA29023.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05451; CAA29024.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05452; CAA29025.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05453; CAA29026.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05454; CAA29027.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05455; CAA29028.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05456; CAA29029.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05457; CAA29030.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05458; CAA29031.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05459; CAA29032.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05460; CAA29033.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05461; CAA29034.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05462; CAA29035.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05463; CAA29036.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05464; CAA29037.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05465; CAA29038.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05466; CAA29039.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05467; CAA29040.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05468; CAA29041.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05469; CAA29042.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05470; CAA29043.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05471; CAA29044.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05472; CAA29045.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05473; CAA29046.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05474; CAA29047.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05475; CAA29048.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05476; CAA29049.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05477; CAA29050.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05478; CAA29051.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05479; CAA29052.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05480; CAA29053.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05481; CAA29054.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05482; CAA29055.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05483; CAA29056.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05484; CAA29057.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05485; CAA29058.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05486; CAA29059.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05487; CAA29060.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05488; CAA29061.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05489; CAA29062.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05490; CAA29063.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05491; CAA29064.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05492; CAA29065.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05493; CAA29066.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05494; CAA29067.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05495; CAA29068.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05496; CAA29069.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05497; CAA29070.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05498; CAA29071.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05499; CAA29072.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05500; CAA29073.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05501; CAA29074.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05502; CAA29075.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05503; CAA29076.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05504; CAA29077.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05505; CAA29078.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05506; CAA29079.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05507; CAA29080.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05508; CAA29081.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05509; CAA29082.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05510; CAA29083.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05511; CAA29084.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05512; CAA29085.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05513; CAA29086.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05514; CAA29087.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05515; CAA29088.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05516; CAA29089.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05517; CAA29090.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05518; CAA29091.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05519; CAA29092.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05520; CAA29093.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05521; CAA29094.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05522; CAA29095.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05523; CAA29096.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05524; CAA29097.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05525; CAA29098.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05526; CAA29099.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05527; CAA29100.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05528; CAA29101.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05529; CAA29102.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05530; CAA29103.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05531; CAA29104.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05532; CAA29105.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05533; CAA29106.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05534; CAA29107.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05535; CAA29108.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05536; CAA29109.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05537; CAA29110.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05538; CAA29111.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05539; CAA29112.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05540; CAA29113.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05541; CAA29114.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05542; CAA29115.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05543; CAA29116.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05544; CAA29117.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05545; CAA29118.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05546; CAA29119.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05547; CAA29120.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05548; CAA29121.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05549; CAA29122.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05550; CAA29123.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05551; CAA29124.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05552; CAA29125.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05553; CAA29126.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05554; CAA29127.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05555; CAA29128.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05556; CAA29129.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05557; CAA29130.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05558; CAA29131.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05559; CAA29132.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05560; CAA29133.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05561; CAA29134.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05562; CAA29135.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05563; CAA29136.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05564; CAA29137.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05565; CAA29138.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05566; CAA29139.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05567; CAA29140.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05568; CAA29141.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05569; CAA29142.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05570; CAA29143.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05571; CAA29144.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05572; CAA29145.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05573; CAA29146.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05574; CAA29147.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05575; CAA29148.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05576; CAA29149.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05577; CAA29150.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05578; CAA29151.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05579; CAA29152.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05580; CAA29153.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05581; CAA29154.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05582; CAA29155.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05583; CAA29156.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05584; CAA29157.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05585; CAA29158.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05586; CAA29159.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05587; CAA29160.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05588; CAA29161.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05589; CAA29162.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05590; CAA29163.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05591; CAA29164.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05592; CAA29165.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05593; CAA29166.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05594; CAA29167.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05595; CAA29168.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05596; CAA29169.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05597; CAA29170.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05598; CAA29171.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05599; CAA29172.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05600; CAA29173.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05601; CAA29174.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05602; CAA29175.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05603; CAA29176.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05604; CAA29177.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05605; CAA29178.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05606; CAA29179.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05607; CAA29180.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05608; CAA29181.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05609; CAA29182.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05610; CAA29183.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05611; CAA29184.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05612; CAA29185.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05613; CAA29186.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05614; CAA29187.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05615; CAA29188.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05616; CAA29189.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05617; CAA29190.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05618; CAA29191.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05619; CAA29192.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05620; CAA29193.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05621; CAA29194.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05622; CAA29195.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05623; CAA29196.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05624; CAA29197.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05625; CAA29198.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05626; CAA29199.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05627; CAA29200.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05628; CAA29201.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05629; CAA29202.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05630; CAA29203.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05631; CAA29204.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05632; CAA29205.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05633; CAA29206.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05634; CAA29207.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05635; CAA29208.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05636; CAA29209.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05637; CAA29210.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05638; CAA29211.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05639; CAA29212.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05640; CAA29213.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05641; CAA29214.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05642; CAA29215.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05643; CAA29216.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05644; CAA29217.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05645; CAA29218.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05646; CAA29219.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05647; CAA29220.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05648; CAA29221.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05649; CAA29222.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05650; CAA29223.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05651; CAA29224.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05652; CAA29225.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05653; CAA29226.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05654; CAA29227.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05655; CAA29228.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05656; CAA29229.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05657; CAA29230.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05658; CAA29231.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05659; CAA29232.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05660; CAA29233.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05661; CAA29234.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05662; CAA29235.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05663; CAA29236.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05664; CAA29237.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05665; CAA29238.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05666; CAA29239.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05667; CAA29240.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05668; CAA29241.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05669; CAA29242.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05670; CAA29243.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05671; CAA29244.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05672; CAA29245.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05673; CAA29246.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05674; CAA29247.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05675; CAA29248.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05676; CAA29249.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05677; CAA29250.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05678; CAA29251.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05679; CAA29252.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05680; CAA29253.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05681; CAA29254.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05682; CAA29255.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05683; CAA29256.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05684; CAA29257.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05685; CAA29258.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05686; CAA29259.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05687; CAA29260.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05688; CAA29261.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05689; CAA29262.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05690; CAA29263.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05691; CAA29264.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05692; CAA29265.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05693; CAA29266.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05694; CAA29267.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05695; CAA29268.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05696; CAA29269.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05697; CAA29270.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05698; CAA29271.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05699; CAA29272.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05700; CAA29273.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05701; CAA29274.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05702; CAA29275.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05703; CAA29276.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05704; CAA29277.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05705; CAA29278.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05706; CAA29279.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05707; CAA29280.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05708; CAA29281.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05709; CAA29282.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05710; CAA29283.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05711; CAA29284.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05712; CAA29285.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05713; CAA29286.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05714; CAA29287.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05715; CAA29288.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05716; CAA29289.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05717; CAA29290.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05718; CAA29291.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05719; CAA29292.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05720; CAA29293.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05721; CAA29294.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05722; CAA29295.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05723; CAA29296.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05724; CAA29297.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05725; CAA29298.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05726; CAA29299.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05727; CAA29300.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05728; CAA29301.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05729; CAA29302.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05730; CAA29303.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05731; CAA29304.1; ALT\_SEQ; Genomic DNA.  
DR EMBL; X05732; CAA29305





Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klauner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 Altschul S.F., Zerborg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Frange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,  
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
 CC Keratin-4 is generally associated with keratin-13.  
 CC -!- TISSUE SPECIFICITY: Expressed in the dorsal and ventral epithelium  
 CC of the tongue. Highest expression levels are detected in the  
 CC suprabasal layer with low levels detected in the basal cell layer.  
 CC Within the suprabasal layer expression is highest in the spinous  
 CC cells, decreases in the granular cells and is not detected in the  
 CC stratum corneum.  
 CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
 CC microfibrillar keratin, I (acidic; 40-55 kDa) [K9 to K20] and II  
 CC (neutral to basic; 56-70 kDa) [K1 to K8].  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: X03491; CAA27207.1; -; mRNA.  
 CC EMBL: BC064008; AAH64008.1; ALT\_INIT; mRNA.  
 CC F01: A23518; A23518.  
 CC HSSP: P08670; 1GK7.  
 CC Ensembl: ENSMUSG00000059668; Mus musculus.  
 CC MGI: MGI:96701; Krt2-4.  
 CC InterPro: IPR001664; IF.  
 CC InterPro: IPR002957; Keratin I.  
 CC InterPro: IPR003054; Keratin II.  
 CC Pfam: PF00038; Filament; 1.  
 CC PRINTS: PR01248; TYPE1KERATIN.  
 CC PRINTS: PR01276; TYPE2KERATIN.  
 CC PROSITE: PS00226; IF; 1.  
 CC PROSITE: PS00226; IF; 1.  
 CC Coiled coil; Intermediate filament; Keratin; Phosphorylation.  
 KW Coiled coil; Intermediate filament; Keratin; Phosphorylation.  
 CHAIN 1 525 /FTid=PRO\_0000063723.  
 FT REGION 1 145 Head.  
 FT REGION 146 455 Rod.  
 FT REGION 146 181 Coiled 1A.  
 FT REGION 182 200 Linker 1.  
 FT REGION 201 292 Coiled 1B.  
 FT REGION 293 316 Linker 12.  
 FT REGION 317 455 Coiled 2.  
 FT REGION 456 524 Tail.  
 FT COMPBIAS 10 117 Gly-rich.  
 FT COMPBIAS 473 510 Gly-rich.  
 FT SITE 397 397 Stutter.  
 FT MOD\_RES 51 51 Phosphoserine (By similarity).  
 FT CONFLICT 13 13 R -> P (in Ref. 1).  
 FT CONFLICT 16 16 S -> T (in Ref. 1).  
 FT CONFLICT 79 79 G -> D (in Ref. 2; AAH64008).  
 FT CONFLICT 388 388 L -> P (in Ref. 1).  
 FT CONFLICT 476 480 ASIGG -> QHWR (in Ref. 1).  
 SQ SEQUENCE 525 AA; 56283 MW; 27A3A3DFE87DF90 CRC64;  
 Query Match 91.1%; Score 41; DB 1; Length 525;  
 Best Local Similarity 88.9%; Pred. No. 15;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AQVEDIAQK 9  
 DB 322 AQVEDIAQK 330  
 RESULT 10  
 Q98UJ3 XENLA PRELIMINARY; PRT; 580 AA.  
 ID Q98UJ3 XENLA PRELIMINARY; PRT; 580 AA.  
 AC Q98UJ3;  
 DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 18.  
 DE Larval keratin XLK.  
 GN Name=xlk;  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus; Xenopus.  
 OC NCBI\_TaxID=8355;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=21240324; PubMed=11342212; DOI=10.1016/S0167-4781(00)00281-5;  
 RA Watanabe Y., Kobayashi H., Suzuki K., Kotani K., Yoshizato K.;  
 RT "New epidermal keratin genes from Xenopus laevis: hormonal and  
 RT regional regulation of their expression during anuran skin  
 RT metamorphosis.";  
 RL Biochim. Biophys. Acta 1517:339-350(2001).  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: AB045599; BAB32829.1; -; mRNA.  
 CC HSSP: P08670; 1GK7.  
 CC GO: GO:0005882; C:intermediate filament; IEA.  
 CC GO: GO:0005198; F:structural molecule activity; IEA.  
 CC InterPro: IPR013326; ApoA/E\_Apolp.  
 CC InterPro: IPR001664; IF.  
 CC InterPro: IPR002957; Keratin I.  
 CC InterPro: IPR003054; Keratin II.  
 CC PANTHER: PTHR18893:SP5; Keratin II; 1.  
 CC Pfam: PF00038; Filament; 1.  
 CC PRINTS: PR01248; TYPE1KERATIN.  
 CC PRINTS: PR01276; TYPE2KERATIN.  
 CC PROSITE: PS00226; IF; 1.  
 CC PROSITE: PS00226; IF; 1.  
 CC Intermediate filament; Keratin.  
 SQ SEQUENCE 580 AA; 59940 MW; 7284092F7B6BC1CB CRC64;  
 Query Match 91.1%; Score 41; DB 2; Length 580;  
 Best Local Similarity 88.9%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AQVEDIAQK 9  
 DB 338 AQVEDIAQK 346  
 RESULT 11  
 Q6EII9 CANFA PRELIMINARY; PRT; 619 AA.  
 ID Q6EII9 CANFA PRELIMINARY; PRT; 619 AA.  
 AC Q6EII9;  
 DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.  
 DT 16-AUG-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 10.  
 DE Epithelial keratin 1.  
 GN Name=ker1;  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 OC Canis.  
 OC NCBI\_TaxID=9615;  
 Query Match 91.1%; Score 41; DB 1; Length 525;  
 Best Local Similarity 88.9%; Pred. No. 15;

```

RN NUCLEOTIDE SEQUENCE.
RX PubMed-15627753; DOI=10.1159/000081527;
RA Credille K.M., Guyon R., Andre C., Murphy K., Tucker K.,
RA Barnhart K.F., Dunstan R.W.;
RT "Comparative sequence analysis and radiation hybrid mapping of two
RT epidermal type II keratin genes in the dog: keratin 1 and keratin
RT 2e.";
RL Cytogenet. Genome Res. 108:328-332(2005).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY318945; AAQ83910.1; -; mRNA.
DR Ensembl; ENSCAFG0000007212; Canis familiaris.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR PANTHER; PTHR18893:SF5; Keratin_II; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 619 AA; 63790 MW; E45A8B709AD0909F CRC64;

Query Match 91.1%; Score 41; DB 2; Length 619;
Best Local Similarity 88.9%; Pred. No. 17;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 357 AQYEDIAQK 365

RESULT 12
QSEI21 CANFA PRELIMINARY; PRT; 633 AA.
AC Q6EIZ1;
DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2004, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Epithelial keratin 2e.
GN Name=K2e;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed-15627753; DOI=10.1159/000081527;
RA Credille K.M., Guyon R., Andre C., Murphy K., Tucker K.,
RA Barnhart K.F., Dunstan R.W.;
RT "Comparative sequence analysis and radiation hybrid mapping of two
RT epidermal type II keratin genes in the dog: keratin 1 and keratin
RT 2e.";
RL Cytogenet. Genome Res. 108:328-332(2005).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY318943; AAQ83908.1; -; mRNA.
DR Ensembl; ENSCAFG0000007223; Canis familiaris.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.

PANTHER: PTHR18893:SF5; Keratin_II; 1.
Pfam: PF00038; Filament; 1.
PRINTS: PR01248; TYPE1KERATIN.
PRINTS: PR01276; TYPE2KERATIN.
PROSITE; PS00226; IF; 1.
Intermediate filament; Keratin.
NON TER 1
SEQUENCE 496 AA; 55152 MW; D327911FBB7A1CBA CRC64;

Query Match 88.9%; Score 40; DB 2; Length 496;
Best Local Similarity 88.9%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 261 AQYEDIAQK 269

RESULT 14
Q6NVR6_XENTR PRELIMINARY; PRT; 508 AA.
ID Q6NVR6_XENTR
```

```

DR PANTHER: PTHR18893:SF5; Keratin_II; 1.
DR Pfam: PF00038; Filament; 1.
DR PRINTS: PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 633 AA; 64567 MW; 445D509166A23CB1 CRC64;

Query Match 91.1%; Score 41; DB 2; Length 633;
Best Local Similarity 88.9%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 366 AQYEDIAQK 374

RESULT 13
O42434 NOTVI PRELIMINARY; PRT; 496 AA.
ID O42434 NOTVI PRELIMINARY; PRT; 496 AA.
AC O42434;
DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.
DT 01-JAN-1998, sequence version 1.
DT 07-FEB-2006, entry version 26.
DE Cytokeratin 8 (Fragment).
OS Notoptthalmus viridescens (Eastern newt) (Triturus viridescens).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
OC Notoptthalmus.
OX NCBI_TaxID=8316;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Limb blastema;
RX MEDLINE=98075904; PubMed=9415422;
RX DOI=10.1002/(SICI)1097-0177(199712)210:4<355::AID-AJAL>3.3.CO;2-K;
RA Corcoran J.P., Ferretti P.;
RT "Keratin 8 and 18 expression in mesenchymal progenitor cells of
RT regenerating limbs is associated with cell proliferation and
RT differentiation.";
RL Dev. Dyn. 210:355-370(1997).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ001296; CAA04656.1; -; mRNA.
DR HSPF; P08670; IGK7.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR18893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
FT NON TER 1
SEQUENCE 496 AA; 55152 MW; D327911FBB7A1CBA CRC64;

Query Match 88.9%; Score 40; DB 2; Length 496;
Best Local Similarity 88.9%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 261 AQYEDIAQK 269

RESULT 14
Q6NVR6_XENTR PRELIMINARY; PRT; 508 AA.
ID Q6NVR6_XENTR
```



AC G6NVR6;  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 16.  
DE Keratin complex 2, basic, gene 5.  
GN Name=krt2-5-prov;  
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodinae; Xenopus; Silurana.  
OX NCBI\_TaxID=8364;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Embryo;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zerbins B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Embryo;  
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT initiative.";  
RL Dev. Dyn. 225:384-391(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Embryo;  
RA Klein S., Gerhard D.S.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; BC067939; AAH67939.1; -; mRNA.  
DR Ensembl; ENSXETG0000020454; Xenopus tropicalis.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin I.  
DR PANTHER; PTHR1893; SFS; Keratin\_II.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 508 AA; 55854 MW; B0FEC774C2F66476 CRC64;

Query Match 88.9%; Score 40; DB 2; Length 508;  
Best Local Similarity 88.9%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQVEDIAQK 9  
DB 277 AQVEDIANK 285  
RESULT 15  
O90ZF7\_RANCA PRELIMINARY; PRT; 511 AA.  
AC O90ZF7;  
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
DT 01-DEC-2001, sequence version 1.  
DT 07-FEB-2006, entry version 17.  
DE Keratin 8.  
GN Name=rk8;  
OS Rana catesbeiana (Bull frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;  
OC Aquarana.  
OX NCBI\_TaxID=8400;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Suzuki K., Sato K., Katsu K., Hayashita H., Kristensen D.B.,  
RA Yoshizato K.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AB056480; BAB63946.1; -; mRNA.  
DR HSSP; P08670; 1GK4  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin I.  
DR PANTHER; PTHR1893; SFS; Keratin\_II.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 511 AA; 56378 MW; 68366DD1EF8EA53D CRC64;  
Query Match 88.9%; Score 40; DB 2; Length 511;  
Best Local Similarity 88.9%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQVEDIAQK 9  
DB 277 AQVEDIANK 285

Search completed: July 5, 2006, 19:18:13  
Job time : 104.625 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:44:12 ; Search time 83.0625 Seconds

(without alignments)  
49.540 Million cell updates/sec

Title: US-10-774-928a-2

Perfect score: 45

Sequence: 1 AQVEDIAQK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_8.\*

1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003Bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*  
10: Geneseq2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	100.0	9	6	AAO26382 Psoriasis
2	45	100.0	9	9	ADV04407 Leishmani
3	45	100.0	644	9	AEAL5501 Human pol
4	42	93.3	420	8	ADP29879 Human sec
5	42	93.3	629	9	AEAL5502 Human pol
6	40	88.9	30	2	AAV14568 Partial a
7	38	84.4	82	4	AAM20776 Peptide #
8	38	84.4	82	4	ABM42330 Peptide #
9	38	84.4	82	4	AAM36136 Peptide #
10	38	84.4	82	4	ABB25821 Protein #
11	38	84.4	82	4	AAM76029 Human bon
12	38	84.4	82	4	AAM63216 Human bra
13	38	84.4	82	4	ABG57755 Human liv
14	38	84.4	82	5	ABG45427 Human pep
15	38	84.4	268	8	ADR98973 Lung spec
16	38	84.4	295	8	ADR98974 Lung spec
17	38	84.4	327	7	ADJ59669 Human hea
18	38	84.4	521	5	ABB78802 Human NOV
19	38	84.4	534	5	ABP43881 57kd kera
20	38	84.4	534	7	ADF09537 Human ker
21	38	84.4	534	9	AEAL5503 Human pol
22	38	84.4	564	5	ABP65226 Hypoxia-r
23	38	84.4	564	8	ADJ75686 Marker ge

24	38	84.4	564	8	ADN04584
25	38	84.4	564	8	ADN04384
26	38	84.4	564	8	ADN04382
27	38	84.4	564	8	ADO55108
28	38	84.4	564	8	ADQ18800
29	38	84.4	564	8	AEAL5493
30	38	84.4	564	9	AEAL5491
31	38	84.4	564	9	AEAL5492
32	38	84.4	564	9	AEAL5494
33	38	84.4	564	9	AEAL5492
34	38	84.4	564	9	AEAL5494
35	38	84.4	564	9	AEAL5494
36	38	84.4	564	9	AEAL5494
37	38	84.4	564	9	AEAL5494
38	38	84.4	564	9	AEAL5494
39	38	84.4	564	9	AEAL5494
40	38	84.4	564	9	AEAL5494
41	38	84.4	564	9	AEAL5494
42	38	84.4	564	9	AEAL5494
43	37	82.2	9	6	ABR75646
44	37	82.2	9	7	ADN07174
45	37	82.2	125	4	ABG20747

#### ALIGNMENTS

RESULT 1

AAO26382

ID AAO26382 standard; peptide; 9 AA.

XX AAO26382;

AC AAO26382;

XX AAO26382;

DT 30-JAN-2003 (first entry)

XX Psoriasis treating immunotherapeutic peptide, SEQ ID NO 2.

DE Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;

XX KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;

XX KW gene therapy.

XX OS Leishmania sp.

XX WO200274239-A2.

XX 26-SEP-2002.

XX 04-MAR-2002; 2002WO-US006496.

XX 16-MAR-2001; 2001US-00809003.

XX (AKIV-) AKIVA LLC.

XX O'daly JA;

XX WPI; 2003-018763/01.

XX New polypeptide and immunogenic variants comprising amino acid sequences

XX of particulate antigens, useful for the treatment and clinical remission

XX of psoriasis.

XX Claim 1; Page 43; 56pp; English.

XX The invention relates to a polypeptide comprising an isolated amino acid

XX sequence or immunogenic variants selected from any of 14 fully defined

XX sequences of 7-16 amino acids, given in the specification. The

XX immunotherapeutic agents and a pharmaceutical compositions comprising

XX polynucleotides and vectors of the invention are useful for the treatment

XX and clinical remission of psoriasis. The isolated nucleic acids are

XX useful as probes. The sequences of the invention can be used in the

XX treatment of disorders by gene therapy. This sequence represents one of

XX the 14 immunotherapeutic peptides of the invention

XX

```
SQ      Sequence 9 AA;
Query Match      100.0%; Score 45; DB 6; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.le+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AQYEDIAQK 9
      |||||
Db      1 AQYEDIAQK 9

RESULT 2
ADV04407
ID      ADV04407 standard; peptide; 9 AA.
XX
AC      ADV04407;
XX
DT      24-FEB-2005 (first entry)
XX
DE      Leishmania immunotherapeutic peptide SEQ ID NO:2.
XX
KW      T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;
KW      psoriasis.
XX
OS      Leishmania sp.
XX
PN      US2004241168-A1.
XX
PD      02-DEC-2004.
XX
PF      09-FEB-2004; 2004US-00774928.
XX
PR      16-MAR-2001; 2001US-00809003.
PR      17-OCT-2003; 2003US-00687892.
XX
PA      (ODAL/) ODALY J A.
PI      Odaly JA;
XX
WPI; 2005-011563/01.
XX
Inhibiting selectively T-cell rolling in human, by administering compound
PT      interfering with cutaneous lymphocyte antigen-E selectin interaction and
PT      leukocyte function associated antigen-1 interaction.
XX
Claim 9; SEQ ID NO 2; 21pp; English.
XX
The invention relates to a novel method for selectively inhibiting T-cell
CC      rolling in a human host, comprising administering a compound that
CC      selectively interferes with the cutaneous lymphocyte antigen (CLA)-E
CC      selectin interaction and leukocyte function associated antigen (LFA)-
CC      1/intercellular adhesion molecule (ICAM) and very late antigen
CC      (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound
CC      includes an immunotherapeutic agent, which comprises a purified protein
CC      extract that is isolated by diethylaminoethyl Sephadex chromatography of
CC      a Nonidet P-40 insoluble particulate antigen fraction derived from
CC      isolated killed cells of anastigotes from one or more species of the
CC      Leishmania genus, where the particulate antigen fraction is solubilized
CC      with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied
CC      to diethylaminoethyl Sephadex and is eluted with a solution comprising
CC      0.1 M sodium chloride, 8 M urea and 0.025 M
CC      Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract
CC      includes polypeptides having apparent molecular weights of 73, 80 and 82
CC      kDa, after total reduction and alkylation. The species is L. amazonensis,
CC      L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the
CC      invention has antipsoriatic and immunostimulant activity, and inhibits T-
CC      cell rolling by interfering with CLA-E selectin interaction and LFA-
CC      1/ICAM and VLA/VCAM interactions. The method is useful for selectively
CC      inhibiting T-cell rolling in a human host, and for treating psoriasis.
CC      The present sequence represents an Leishmania peptide used in the
CC      invention.
XX
SQ      Sequence 9 AA;
Query Match      100.0%; Score 45; DB 9; Length 644;
Best Local Similarity 100.0%; Pred. No. 3.5;

Query Match      100.0%; Score 45; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.le+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AQYEDIAQK 9
      |||||
Db      1 AQYEDIAQK 9

RESULT 3
AEAL5501
ID      AEAL5501 standard; protein; 644 AA.
XX
AC      AEAL5501;
XX
DT      28-JUL-2005 (first entry)
XX
DE      Human polypeptide #120.
XX
KW      Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;
KW      colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;
KW      cytostatic; neoplasm.
XX
OS      Homo sapiens.
XX
PN      WO2005047534-A2.
XX
PD      26-MAY-2005.
XX
PF      15-OCT-2004; 2004WO-EP011599.
XX
PR      28-OCT-2003; 2003EP-00024565.
XX
PA      (FARB ) BAYER HEALTHCARE AG.
PI      Wirtz R, Munnes M;
XX
WPI; 2005-372393/38.
DR      N-PSDB; AEAL5423.
XX
Predicting a response to cancer treatment by detecting at least 2
PT      markers, which are genes or genomic nucleic acid sequences that are
PT      located on one chromosomal region, which is altered in malignant
PT      neoplasia.
XX
Claim 7; SEQ ID NO 463; 464pp; English.
XX
The invention relates to a method of predicting response to cancer
CC      treatment comprising detection of at least 2 markers, where the markers
CC      are genes and fragments or genomic nucleic acid sequences that are
CC      located on one chromosomal region, which is altered in malignant
CC      neoplasia. The invention also relates to a method for the prediction,
CC      diagnosis or prognosis of malignant neoplasia, methods for detecting
CC      deregulations in malignant neoplasia and breast cancer, a method of
CC      determining the phenotype of a cell or tissue, a method for identifying
CC      genomic regions which are altered on the chromosomal level and encode
CC      genes that are linked by function and are differentially expressed in
CC      malignant neoplasia and breast cancer, methods of screening for agents
CC      which regulate the activity of a polypeptide or a polynucleotide and
CC      antibodies that specifically bind to a full length or partial
CC      polypeptide. The method is useful for predicting response to cancer
CC      treatment. The methods and compositions are useful for predicting,
CC      diagnosing, prognosing, preventing or treating malignant neoplasia
CC      including breast cancer, ovarian cancer, gastric cancer, colon cancer,
CC      esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell
CC      lung cancer. This sequence represents a human polypeptide used in the
CC      scope of the invention.
XX
SQ      Sequence 644 AA;
```

```
Matches      9;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy      1 AQVEDIAQK 9
Db      356 AQVEDIAQK 364

RESULT 4
ADP29879
ID   ADP29879 standard; protein; 420 AA.
XX
AC   ADP29879;
DT   01-DEC-2005 (revised)
DT   12-AUG-2004 (first entry)
XX
DE   Human secreted protein SEQ ID #1877.
XX
KW   Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
KW   cancer; inflammatory; immune; human secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO2004035732-A2.
XX
PD   29-APR-2004.
XX
PF   28-AUG-2003; 2003WO-US026780.
XX
PR   29-AUG-2002; 2002US-0406576P.
PR   29-AUG-2002; 2002US-0406579P.
PR   29-AUG-2002; 2002US-0406585P.
PR   29-AUG-2002; 2002US-0406588P.
PR   29-AUG-2002; 2002US-0406608P.
PR   29-AUG-2002; 2002US-0406611P.
PR   29-AUG-2002; 2002US-0406612P.
PR   29-AUG-2002; 2002US-0406616P.
PR   29-AUG-2002; 2002US-0406640P.
PR   29-AUG-2002; 2002US-0406642P.
PR   29-AUG-2002; 2002US-0406646P.
PR   29-AUG-2002; 2002US-0406653P.
PR   29-AUG-2002; 2002US-0406655P.
PR   29-AUG-2002; 2002US-0406666P.
PR   17-SEP-2002; 2002US-0410946P.
PR   17-SEP-2002; 2002US-0410947P.
PR   17-SEP-2002; 2002US-0410948P.
PR   17-SEP-2002; 2002US-0410949P.
PR   17-SEP-2002; 2002US-0410953P.
PR   17-SEP-2002; 2002US-0410957P.
PR   17-SEP-2002; 2002US-0410958P.
PR   17-SEP-2002; 2002US-0410959P.
PR   17-SEP-2002; 2002US-0410960P.
PR   17-SEP-2002; 2002US-0410961P.
PR   17-SEP-2002; 2002US-0410962P.
PR   17-SEP-2002; 2002US-0411019P.
PR   17-SEP-2002; 2002US-0411022P.
PR   17-SEP-2002; 2002US-0411023P.
PR   17-SEP-2002; 2002US-0411024P.
PR   17-SEP-2002; 2002US-0411032P.
PR   17-SEP-2002; 2002US-0411035P.
PR   17-SEP-2002; 2002US-0411037P.
PR   17-SEP-2002; 2002US-0411041P.
PR   17-SEP-2002; 2002US-0411045P.
PR   17-SEP-2002; 2002US-0411046P.
PR   17-SEP-2002; 2002US-0411048P.
PR   17-SEP-2002; 2002US-0411052P.
PR   17-SEP-2002; 2002US-0411055P.
PR   17-SEP-2002; 2002US-0411073P.
PR   17-SEP-2002; 2002US-0411082P.
PR   17-SEP-2002; 2002US-0411101P.
PR   17-SEP-2002; 2002US-0411111P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.

18-APR-2003; 2003US-0463716P.
18-APR-2003; 2003US-0463732P.
02-MAY-2003; 2003US-0467199P.
02-MAY-2003; 2003US-0467201P.
02-MAY-2003; 2003US-0467203P.
02-MAY-2003; 2003US-0467230P.
19-MAY-2003; 2003US-0471306P.
19-MAY-2003; 2003US-0471336P.
22-MAY-2003; 2003US-0472420P.
09-JUN-2003; 2003US-0472430P.
09-JUN-2003; 2003US-0476609P.
08-JUL-2003; 2003US-0485218P.
08-JUL-2003; 2003US-0485223P.
08-JUL-2003; 2003US-0485224P.
14-JUL-2003; 2003US-0486446P.
14-JUL-2003; 2003US-0486480P.
15-JUL-2003; 2003US-0486891P.
15-JUL-2003; 2003US-0486960P.
08-AUG-2003; 2003US-0493341P.
08-AUG-2003; 2003US-0493370P.
08-AUG-2003; 2003US-0493573P.
08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI   Halenbeck RF, Huang MM, Kothakota S, Haishan L, Linnemann T;
PI   Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;
XX
DR   WPI; 2004-348438/32.
XX
PT   New nucleic acid molecule for diagnosing, preventing or treating diseases
PT   such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
PT   genetic, bacterial and viral diseases.
XX
PS   Claim 1; SEQ ID NO 1877; 428pp; English.
XX
CC   The present invention relates to an isolated nucleic acid molecule
CC   encoding a polypeptide which is believed to be cytostatic,
CC   antiinflammatory, immunosuppressive, antibacterial and virucidal. The
CC   composition and methods are useful for diagnosing, preventing and
CC   treating diseases such as proliferative (e.g. cancer), inflammatory,
CC   immune, metabolic, genetic, bacterial and viral diseases. The present
CC   sequence represents a human secreted protein. The present sequence is
CC   available on WIPWEB and is not in the specification.
CC
CC   Revised record issued on 01-DEC-2005 : Sequence description line
CC   corrected
XX
SQ   Sequence 420 AA;

Query Match      93.3%; Score 42; DB 8; Length 420;
Best Local Similarity 88.9%; Pred. No. 8.7;
Matches      8; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

QY      1 AQVEDIAQK 9
Db      271 AQVEDIAQK 279

RESULT 5
AEA15502
ID   AEA15502 standard; protein; 629 AA.
XX
AC   AEA15502;
XX
DT   28-JUL-2005 (first entry)
XX
DE   Human polypeptide #121.
XX
KW   Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;
```

KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
XX cytostatic; neoplasm.  
OS Homo sapiens.  
XX WO2005047534-A2.  
PN 26-MAY-2005.  
XX 15-OCT-2004; 2004WO-EP011599.  
PF 28-OCT-2003; 2003EP-00024565.  
XX (FARB ) BAYER HEALTHCARE AG.  
XX Wirtz R, Munnes M;  
PI WPI; 2005-372393/38.  
DR N-PSDB; AEA15424.  
XX Predicting a response to cancer treatment by detecting at least 2  
PT markers, which are genes or genomic nucleic acid sequences that are  
PT located on one chromosomal region, which is altered in malignant  
PT neoplasia.  
XX Claim 7; SEQ ID NO 464; 464pp; English.  
XX The invention relates to a method of predicting response to cancer  
CC treatment comprising detection of at least 2 markers, where the markers  
CC are genes and fragments or genomic nucleic acid sequences that are  
CC located on one chromosomal region, which is altered in malignant  
CC neoplasia. The invention also relates to a method for the prediction,  
CC diagnosis or prognosis of malignant neoplasia, methods for detecting  
CC deregulations in malignant neoplasia and breast cancer, a method of  
CC determining the phenotype of a cell or tissue, a method for identifying  
CC genomic regions which are altered on the chromosomal level and encode  
CC genes that are linked by function and are differentially expressed in  
CC malignant neoplasia and breast cancer. Methods of screening for agents  
CC which regulate the activity of a polypeptide or a polynucleotide and  
CC antibodies that specifically bind to a full length or partial  
CC polypeptide. The method is useful for predicting response to cancer  
CC treatment. The methods and compositions are useful for predicting,  
CC diagnosing, prognosing, preventing or treating malignant neoplasia  
CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
CC lung cancer. This sequence represents a human polypeptide used in the  
XX scope of the invention.  
SQ Sequence 629 AA;  
Query Match 93.3%; Score 42; DB 9; Length 629;  
Best Local Similarity 88.9%; Pred. No. 14; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 1;  
QY 1 AQYEDIAQK 9  
Db 376 AQYEDIAQR 384  
RESULT 6  
AAV14568  
ID AAV14568 standard; peptide; 30 AA.  
XX AC AAV14568;  
XX 14-SEP-1999 (first entry)  
XX Partial amino acid sequence of D-sequence binding protein.  
DE D-sequence; binding protein; phosphorylation; adeno-associated virus;  
KW AAV; transgene; expression; DNA synthesis; cancer; EMSA; detection;  
KW competitive electrophoretic mobility shift assay.  
XX

OS Synthetic.  
XX WO9911652-A1.  
XX 11-MAR-1999.  
XX 01-SEP-1998; 98WO-US018194.  
XX 02-SEP-1997; 97US-0056052P.  
XX (ADRE-) ADVANCED RES & TECHNOLOGY INST.  
XX Srivastava A, Qing K, Wang X, Ponnazhagan S, Bajpai A;  
XX WPI; 1999-394757/33.  
XX New isolated D sequence binding protein.  
XX Disclosure; Page 23; 163pp; English.  
XX The invention relates to an isolated and purified D sequence binding  
CC protein (D-BP) having a molecular weight of about 53 kD. This peptide  
CC represents a partial amino acid sequence of the complete D-BP. The D-BP  
CC is phosphorylated at tyrosine residues and blocks AAV-mediated transgene  
CC expression in infected cells by inhibiting the leading strand viral DNA  
CC synthesis. Inhibiting the function of D-BP in host cells can increase  
CC expression of selected nucleic acids from an AAV, particularly for  
CC increasing therapeutic nucleic acids for treating a disease such as  
CC cancer  
SQ Sequence 30 AA;  
Query Match 88.9%; Score 40; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. No. 1-2; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0;  
QY 1 AQYEDIAQ 8  
Db 17 AQYEDIAQ 24  
RESULT 7  
AAM20776  
ID AAM20776 standard; protein; 82 AA.  
XX AC AAM20776;  
XX 12-OCT-2001 (first entry)  
XX Peptide #7210 encoded by probe for measuring cervical gene expression.  
DE Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.  
XX Homo sapiens.  
XX WO200157278-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000670.  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI

```
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human cervical epithelial cells.
XX PS Claim 27; SEQ ID NO 25602; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX CC can be used to produce a single exon microarray, which can be used for
XX CC measuring human gene expression in a sample derived from human cervical
XX CC epithelial cells. By measuring gene expression, the probes are therefore
XX CC useful in grading and/or staging of diseases of the cervix, notably
XX CC cervical cancer. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 82 AA;
XX Query Match 84.4%; Score 38; DB 4; Length 82;
XX Best Local Similarity 77.8%; Pred. No. 9.3;
XX Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AQVEDIAQK 9
XX DB ||||:||||:
XX 35 AQVEEIAQR 43
XX
XX RESULT 8
XX ID ABB42330 standard; peptide; 82 AA.
XX AC ABB42330;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #9836 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human foetal liver.
XX PS Claim 27; SEQ ID NO 34965; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human foetal liver. The
```

```
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 82 AA;
XX Query Match 84.4%; Score 38; DB 4; Length 82;
XX Best Local Similarity 77.8%; Pred. No. 9.3;
XX Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AQVEDIAQK 9
XX DB ||||:||||:
XX 35 AQVEEIAQR 43
XX
XX RESULT 9
XX ID AAM36136 standard; protein; 82 AA.
XX AC AAM36136;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #10173 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 36405; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders
XX SQ Sequence 82 AA;
XX Query Match 84.4%; Score 38; DB 4; Length 82;
XX Best Local Similarity 77.8%; Pred. No. 9.3;
XX Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AQVEDIAQK 9
XX DB ||||:||||:
XX 35 AQVEEIAQR 43
```

```

RESULT 10
ABB25821
ID ABB25821 standard; protein; 82 AA.
XX AC ABB25821;
XX DT 23-JAN-2002 (first entry)
XX DE Protein #7820 encoded by probe for measuring heart cell gene expression.
XX KW Human; gene expression; heart; microarray; vascular system;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488990/53.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 15; SEQ ID NO 27591; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart (see
XX CC ABA21535-ABA41303). The present sequence is a protein encoded by one such
XX CC probe. The probes may be used for predicting, measuring and displaying
XX CC gene expression in samples derived from the human heart via microarrays.
XX CC By measuring gene expression, the probes are useful for predicting,
XX CC diagnosing, grading, staging, monitoring and prognosing diseases of the
XX CC human heart and vascular system e.g. cardiovascular disease,
XX CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 82 AA;

Query Match 84.4%; Score 38; DB 4; Length 82;
Best Local Similarity 77.8%; Pred. No. 9.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDIAQK 9
Db 35 AQYEEIAQR 43
|||||:
35 AQYEEIAQR 43

RESULT 11
AAM76029
ID AAM76029 standard; protein; 82 AA.
XX AC AAM76029;
XX DT 06-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 35321.
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000667.
XX PR 04-FEB-2000; 2000US-0180312P.

```

```

XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 36335.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488990/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO 36335; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention
XX SQ Sequence 82 AA;

Query Match 84.4%; Score 38; DB 4; Length 82;
Best Local Similarity 77.8%; Pred. No. 9.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDIAQK 9
Db 35 AQYEEIAQR 43
|||||:
35 AQYEEIAQR 43

RESULT 12
AAM63216
ID AAM63216 standard; protein; 82 AA.
XX AC AAM63216;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 35321.
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000667.
XX PR 04-FEB-2000; 2000US-0180312P.

```



PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains.  
PT Example 4; SEQ ID NO 35321; 650pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention  
XX  
SQ Sequence 82 AA;  
Query Match 84.4%; Score 38; DB 4; Length 82;  
Best Local Similarity 77.8%; Pred. No. 9.3;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AQVEDIAQK 9  
Db 35 AQVEEIAQR 43  
RESULT 13  
ID ABG57755 standard; peptide; 82 AA.  
AC ABG57755;  
XX  
DT 25-FEB-2003 (first entry)  
XX Human liver peptide, SEQ ID No 36403.  
DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.  
KW Homo sapiens.  
OS  
XX  
PN WO200157273-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000664.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488898/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human adult liver.  
XX Claim 27; SEQ ID NO 36403; 658pp; English.  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABG47348-ABG5930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 82 AA;  
Query Match 84.4%; Score 38; DB 4; Length 82;  
Best Local Similarity 77.8%; Pred. No. 9.3;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AQVEDIAQK 9  
Db 35 AQVEEIAQR 43  
RESULT 14  
ID ABG45427 standard; peptide; 82 AA.  
XX  
AC ABG45427;  
XX  
DT 19-AUG-2002 (first entry)  
XX Human peptide encoded by genome-derived single exon probe SEQ ID 35092.  
DE Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000665.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:52:16 ; Search time 47.625 Seconds  
(without alignments)  
87.537 Million cell updates/sec

Title: US-10-774-928A-2

Perfect score: 45

Sequence: 1 AQVEDIAQK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	9	5	US-10-774-928-2
2	41	91.1	524	3	US-09-976-782-88
3	38	84.4	82	3	US-09-864-761-41119
4	38	84.4	327	4	US-10-408-765A-1475
5	38	84.4	521	3	US-09-976-782-10
6	38	84.4	534	3	US-09-976-782-86
7	38	84.4	534	3	US-09-976-782-87
8	38	84.4	534	4	US-10-231-913-86
9	38	84.4	534	4	US-10-231-913-87
10	38	84.4	534	5	US-10-367-057-38
11	38	84.4	564	4	US-10-170-385-421
12	38	84.4	564	5	US-10-723-860-1619
13	38	84.4	564	4	US-10-756-149-5130
14	38	84.4	564	5	US-10-631-467-938
15	38	84.4	564	6	US-11-037-713-27
16	38	84.4	585	5	US-10-821-234-875
17	38	84.4	608	4	US-10-094-749-2446
18	38	84.4	645	3	US-09-919-172-41
19	38	84.4	645	5	US-10-752-986-41
20	37	82.2	9	4	US-10-285-394-33
21	37	82.2	125	5	US-10-450-763-51106
22	37	82.2	181	3	US-09-864-761-35721
23	37	82.2	222	4	US-10-231-913-263
24	37	82.2	267	5	US-10-450-763-38184
25	37	82.2	273	5	US-10-450-763-56735
26	37	82.2	275	5	US-10-450-763-45641
27	37	82.2	279	6	US-11-013-684-21

28 37 82.2 310 6 US-11-156-300-71 Sequence 71, Appl  
29 37 82.2 311 6 US-11-156-300-70 Sequence 70, Appl  
30 37 82.2 312 3 US-09-976-782-32 Sequence 32, Appl  
31 37 82.2 312 4 US-10-080-334-167 Sequence 167, Appl  
32 37 82.2 312 4 US-10-080-334-188 Sequence 188, Appl  
33 37 82.2 312 4 US-10-080-334-270 Sequence 270, Appl  
34 37 82.2 312 4 US-10-231-913-260 Sequence 260, Appl  
35 37 82.2 315 4 US-10-231-913-259 Sequence 259, Appl  
36 37 82.2 318 4 US-10-231-913-261 Sequence 261, Appl  
37 82.2 398 5 US-10-883-020-5 Sequence 5, Appl  
38 37 82.2 422 3 US-09-779-307-18 Sequence 18, Appl  
39 37 82.2 424 5 US-10-450-763-49620 Sequence 49620, A  
40 37 82.2 428 3 US-09-779-307-21 Sequence 21, Appl  
41 37 82.2 430 3 US-09-779-307-19 Sequence 19, Appl  
42 37 82.2 482 3 US-09-779-307-17 Sequence 17, Appl  
43 37 82.2 482 4 US-10-080-334-187 Sequence 187, Appl  
44 37 82.2 482 5 US-10-883-020-1 Sequence 1, Appl  
45 37 82.2 483 3 US-09-919-497-79 Sequence 79, Appl

#### ALIGNMENTS

##### RESULT 1

US-10-774-928-2  
; Sequence 2, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-2

Query Match 100.0%; Score 45; DB 5; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.9e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9  
Db 1 AQVEDIAQK 9

##### RESULT 2

US-09-976-782-88  
; Sequence 88, Application US/09976782  
; Publication No. US20030190715A1  
; GENERAL INFORMATION:  
; APPLICANT: Grosse et al  
; TITLE OF INVENTION: No. US20030190715A1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-157  
; CURRENT APPLICATION NUMBER: US/09/976,782  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,113  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: 60/240,662  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,732  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,625  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,703  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/241,190  
; PRIOR FILING DATE: 2000-10-16

; PRIOR APPLICATION NUMBER: 60/240,637  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,669  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/262,455  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/240,648  
; PRIOR FILING DATE: 2000-10-16  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 88  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-782-88

Query Match 91.1%; Score 41; DB 3; Length 524;  
Best Local Similarity 88.9%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDIAQK 9  
Db 322 AQYEDIAQK 330

## RESULT 3

US-09-864-761-41119  
; Sequence 41119, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 41119  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC024196.2  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.6  
; OTHER INFORMATION: EST HUMAN HIT: BE272256.1, EVALUE 5.00e-17  
; OTHER INFORMATION: SWISSPROT HIT: P35908, EVALUE 2.00e-24  
US-09-864-761-41119

Query Match 84.4%; Score 38; DB 3; Length 82;  
Best Local Similarity 77.8%; Pred. No. 12;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDIAQK 9  
Db 35 AQYEDIAQK 43

## RESULT 4

US-10-408-765A-1475  
; Sequence 1475, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Faby, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; FILE REFERENCE: 660088.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1475  
; LENGTH: 327  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-1475

Query Match 84.4%; Score 38; DB 4; Length 327;  
Best Local Similarity 77.8%; Pred. No. 51;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDIAQK 9  
Db 120 AQYEDIAQK 128

## RESULT 5

US-09-976-782-10  
; Sequence 10, Application US/09976782  
; Publication No. US20030190715A1  
; GENERAL INFORMATION:  
; APPLICANT: Grose et al  
; TITLE OF INVENTION: No. US20030190715A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-157

; CURRENT APPLICATION NUMBER: US/09/976,782  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,113  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: 60/240,662  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,732  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,625  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,703  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/241,190  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,637  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,669  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/262,455  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/240,648  
; PRIOR FILING DATE: 2000-10-16  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 521  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-782-10

Query Match 84.4%; Score 38; DB 3; Length 521;  
Best Local Similarity 77.8%; Pred. No. 83;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9  
Db 314 AQVEEIAQR 322  
|||||:

## RESULT 6

US-09-976-782-86  
; Sequence 86, Application US/09976782  
; Publication No. US20030190715A1  
; GENERAL INFORMATION:  
; APPLICANT: Grosse et al

; TITLE OF INVENTION: No. US20030190715A1e1 Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-157  
; CURRENT APPLICATION NUMBER: US/09/976,782  
; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: 60/240,113  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: 60/240,662  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,732  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,625  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,703  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/241,190  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,637  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,669  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/262,455  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/240,648  
; PRIOR FILING DATE: 2000-10-16  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 86  
; LENGTH: 534

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-782-86

Query Match 84.4%; Score 38; DB 3; Length 534;  
Best Local Similarity 77.8%; Pred. No. 85;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9  
Db 327 AQVEEIAQR 335  
|||||:

## RESULT 7

US-09-976-782-87  
; Sequence 87, Application US/09976782  
; Publication No. US20030190715A1  
; GENERAL INFORMATION:  
; APPLICANT: Grosse et al

; TITLE OF INVENTION: No. US20030190715A1e1 Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-157  
; CURRENT APPLICATION NUMBER: US/09/976,782  
; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: 60/240,113  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: 60/240,662  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,732  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,625  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,703  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/241,190  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,637  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,669  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/262,455  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/240,648  
; PRIOR FILING DATE: 2000-10-16  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 87  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-782-87

Query Match 84.4%; Score 38; DB 3; Length 534;  
Best Local Similarity 77.8%; Pred. No. 85;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9  
Db 327 AQVEEIAQR 335  
|||||:

## RESULT 8

US-10-231-913-86  
; Sequence 86, Application US/10231913  
; Publication No. US20040005576A1  
; GENERAL INFORMATION:  
; APPLICANT: Guo, Xiaojia S.

; APPLICANT: Li, Li  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.

```
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schalomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-913-86

Query Match      84.4%; Score 38; DB 4; Length 534;
Best Local Similarity 77.8%; Pred. No. 85;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AQYEDIAQK 9
Db      327 AQYEEIAQR 335

RESULT 9
US-10-231-913-87
; Sequence 87, Application US/10231913
; Publication No. US2004000576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schalomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
```

```
; APPLICANT: Boldog, Ference L.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-913-87

Query Match      84.4%; Score 38; DB 4; Length 534;
Best Local Similarity 77.8%; Pred. No. 85;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AQYEDIAQK 9
Db      327 AQYEEIAQR 335

RESULT 10
US-10-367-057-38
; Sequence 38, Application US/10367057
; Publication No. US20050100554A1
; GENERAL INFORMATION:
; APPLICANT: Cuthill, Scott;
; APPLICANT: Jackson, Amanda;
; APPLICANT: Lewin, David A.;
; APPLICANT: Ooi, Chean Eng
; FILE OF INVENTION: Complexes and Methods of Using Same
; FILE REFERENCE: 21402-559
; CURRENT APPLICATION NUMBER: US/10/367,057
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: 60/256,911
; PRIOR FILING DATE: 2002-02-14
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 38
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-367-057-38

Query Match      84.4%; Score 38; DB 5; Length 534;
Best Local Similarity 77.8%; Pred. No. 85;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AQYEDIAQK 9
Db      327 AQYEEIAQR 335
```

Db 327 AQYEETIAQR 335

RESULT 11  
US-10-170-385-421  
; Sequence 421, Application US/10170385  
; Publication No. US20030203372A1  
; GENERAL INFORMATION:  
; APPLICANT: Ward, Neil Raymond  
; APPLICANT: Mundy, Christopher Robert  
; APPLICANT: Kan, On  
; APPLICANT: Harris, Robert Alan  
; APPLICANT: White, Jonathan  
; APPLICANT: Binley, Katie Mary  
; APPLICANT: Rayner, William Nigel  
; APPLICANT: Naylor, Stuart  
; APPLICANT: Kingsman, Susan Mary  
; APPLICANT: Krige, David  
; TITLE OF INVENTION: ANALYSIS METHOD  
; FILE REFERENCE: 532682000100  
; CURRENT APPLICATION NUMBER: US/10/170,385  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: PCT/GB02/01662  
; PRIOR FILING DATE: 2002-04-08  
; PRIOR APPLICATION NUMBER: PCT/GB01/05458  
; PRIOR FILING DATE: 2001-12-10  
; NUMBER OF SEQ ID NOS: 549  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 421  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-170-385-421

Query Match 84.4%; Score 38; DB 4; Length 564;  
Best Local Similarity 77.8%; Pred. No. 90;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
|||||:  
Db 339 AQYEETIAQR 347

RESULT 12  
US-10-723-860-1619  
; Sequence 1619, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882,0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1619  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-723-860-1619

Query Match 84.4%; Score 38; DB 5; Length 564;  
Best Local Similarity 77.8%; Pred. No. 90;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
|||||:  
Db 339 AQYEETIAQR 347

RESULT 13

US-10-756-149-5130  
; Sequence 5130, Application US/10756149  
; Publication No. US20050181375A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER  
; FILE REFERENCE: file  
; CURRENT APPLICATION NUMBER: US/10/756,149  
; CURRENT FILING DATE: 2004-01-12  
; NUMBER OF SEQ ID NOS: 5818  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5130  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-756-149-5130

Query Match 84.4%; Score 38; DB 5; Length 564;  
Best Local Similarity 77.8%; Pred. No. 90;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
|||||:  
Db 339 AQYEETIAQR 347

RESULT 14

US-10-631-467-938  
; Sequence 938, Application US/10631467  
; Publication No. US20050208496A1  
; GENERAL INFORMATION:  
; APPLICANT: Genox Research Inc.  
; TITLE OF INVENTION: Method for testing for bronchial asthma, or chronic obstructive p  
; TITLE OF INVENTION: disease  
; FILE REFERENCE: 3462.1005-000  
; CURRENT APPLICATION NUMBER: US/10/631,467  
; CURRENT FILING DATE: 2003-07-31  
; PRIOR APPLICATION NUMBER: JP 2003-077212  
; PRIOR FILING DATE: 2003-03-20  
; PRIOR APPLICATION NUMBER: JP 2002-229312  
; PRIOR FILING DATE: 2002-08-06  
; NUMBER OF SEQ ID NOS: 2086  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 938  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-631-467-938

Query Match 84.4%; Score 38; DB 5; Length 564;  
Best Local Similarity 77.8%; Pred. No. 90;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
|||||:  
Db 339 AQYEETIAQR 347

RESULT 15

US-11-037-713-27  
; Sequence 27, Application US/11037713  
; Publication No. US20050221398A1  
; GENERAL INFORMATION:  
; APPLICANT: JACQUEMIER, JOCELYNE  
; APPLICANT: BERTUCCI, FRANCOIS  
; APPLICANT: BIRNBAUM, DANIEL  
; APPLICANT: DEBONO, STEPHANE  
; APPLICANT: TAGETT, REBECCA

```

; TITLE OF INVENTION: PROTEIN EXPRESSION PROFILING AND BREAST CANCER
; FILE REFERENCE: 1016-R-04 (B)
; CURRENT APPLICATION NUMBER: US/11/037,713
; CURRENT FILING DATE: 2005-01-18
; PRIOR APPLICATION NUMBER: 60/537,412
; PRIOR FILING DATE: 2004-01-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 27
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-037-713-27

```

```

Query Match      84.4%; Score 38; DB 6; Length 564;
Best Local Similarity 77.8%; Pred. No. 90;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AQYEDIAQK 9
Db      339 AQYEEIAQR 347

```

Search completed: July 5, 2006, 20:57:18  
Job time : 47.625 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:53:13 ; Search time 4.125 Seconds  
(without alignments)  
58.543 Million cell updates/sec

Title: US-10-774-928A-2

Perfect score: 45

Sequence: 1 AQVEDIAQK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.\*

- 1: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /EMC\_Celerra\_SID33/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	37	82.2	483	7	US-11-105-233-180
2	35	77.8	655	6	Sequence 180, App
3	35	77.8	656	6	Sequence 51488, A
4	33	73.3	915	6	Sequence 53558, A
5	32	71.1	430	7	Sequence 45529, A
6	32	71.1	450	7	Sequence 181, App
7	32	71.1	450	7	Sequence 4090, App
8	31	68.9	175	6	Sequence 284, App
9	31	68.9	195	6	Sequence 4732, App
10	31	68.9	195	6	Sequence 4731, App
11	31	68.9	195	6	Sequence 36, App
12	31	68.9	196	6	Sequence 345, App
13	31	68.9	196	6	Sequence 8, App
14	31	68.9	196	7	Sequence 16, App
15	31	68.9	196	7	Sequence 304, App
16	31	68.9	197	6	Sequence 407, App
17	31	68.9	197	6	Sequence 66, App
18	31	68.9	197	6	Sequence 294, App
19	31	68.9	197	6	Sequence 1676, App
20	30	66.7	103	6	Sequence 325, App
21	30	66.7	175	6	Sequence 28513, A
22	30	66.7	196	6	Sequence 1173, App
23	30	66.7	196	6	Sequence 1172, App
24	30	66.7	197	6	Sequence 136, App
25	30	66.7	197	6	Sequence 20, App
					Sequence 22, App

```

26 30 66.7 197 7 US-11-251-208-417 Sequence 417, App
27 30 66.7 197 7 US-11-251-208-431 Sequence 431, App
28 30 66.7 198 6 US-10-566-644-62 Sequence 62, Appl
29 30 66.7 220 6 US-10-449-902-32059 Sequence 32059, A
30 30 66.7 220 6 US-10-449-902-47373 Sequence 47373, A
31 30 66.7 220 6 US-10-449-902-56490 Sequence 56490, A
32 30 66.7 275 6 US-10-449-902-48344 Sequence 48344, A
33 30 66.7 282 6 US-10-953-349-21448 Sequence 21448, A
34 30 66.7 329 6 US-10-471-571A-4880 Sequence 4880, Ap
35 30 66.7 394 6 US-10-449-902-35309 Sequence 35309, A
36 30 66.7 490 7 US-11-293-697-4335 Sequence 4335, Ap
37 29 64.4 130 6 US-10-953-349-525 Sequence 525, App
38 29 64.4 143 6 US-10-449-902-53316 Sequence 53316, A
39 29 64.4 290 6 US-10-449-902-54864 Sequence 54864, A
40 29 64.4 469 6 US-10-505-928-457 Sequence 457, App
41 29 64.4 550 6 US-10-449-902-29669 Sequence 29669, A
42 29 64.4 574 6 US-10-449-902-52042 Sequence 52042, A
43 29 64.4 591 6 US-10-449-902-53151 Sequence 53151, A
44 29 64.4 601 7 US-11-293-697-3847 Sequence 3847, Ap
45 29 64.4 603 6 US-10-471-571A-3718 Sequence 3718, Ap

```

#### ALIGNMENTS

```

RESULT 1
US-11-105-233-180
; Sequence 180, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; FILE OF INVENTION: Tumors
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-180

```

```

Query Match      82.2%; Score 37; DB 7; Length 483;
Best Local Similarity 77.8%; Pred. No. 4.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 AQVEDIAQK 9
Db 265 AQVEDIANR 273

```

```

RESULT 2
US-10-449-902-51488
; Sequence 51488, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51488

```

```
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-51488

Query Match      77.8%; Score 35; DB 6; Length 655;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 AQYEDIAQK 9
Db 171 AQNEDVAQK 179

RESULT 3
US-10-449-902-53558
; Sequence 53558, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53558
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53558

Query Match      77.8%; Score 35; DB 6; Length 656;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 AQYEDIAQK 9
Db 172 AQNEDVAQK 180

RESULT 4
US-10-449-902-45529
; Sequence 45529, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45529
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-45529

Query Match      73.3%; Score 33; DB 6; Length 915;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 898 SQYEDIYQR 906

RESULT 5
US-11-105-233-181
; Sequence 181, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-181

Query Match      71.1%; Score 32; DB 7; Length 430;
Best Local Similarity 55.6%; Pred. No. 40;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 254 AQYDELARK 262

RESULT 6
US-11-293-697-4090
; Sequence 4090, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4090
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-4090

Query Match      71.1%; Score 32; DB 7; Length 450;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 259 AQYETIAAK 267

RESULT 7
US-11-251-208-284
; Sequence 284, Application US/11251208
; Publication No. US20060137043A1
; GENERAL INFORMATION:
; APPLICANT: Puzio, Piotr
; APPLICANT: Chardonnens, Agnes
```

```
; APPLICANT: Shirley, Amber
; APPLICANT: Wang, Xi-Qing
; APPLICANT: Sarria-Millan, Rodrigo
; APPLICANT: McKersie, Bryan
; APPLICANT: Chen, Ruoying
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING PROTEINS ASSOCIATED WITH ABIOTIC
; TITLE OF INVENTION: STRESS RESPONSE AND PLANT CELLS AND PLANTS WITH INCREASED
; TITLE OF INVENTION: TOLERANCE TO ENVIRONMENTAL STRESS
; FILE REFERENCE: 13311-00015-US
; CURRENT APPLICATION NUMBER: US/11/251.208
; CURRENT FILING DATE: 2005-10-14
; PRIOR FILING DATE: 2004-04-15
; PRIOR APPLICATION NUMBER: PCT/US2004/011888
; PRIOR FILING DATE: 2004-04-15
; PRIOR APPLICATION NUMBER: EP 03008080.8
; PRIOR FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: EP 03009728.1
; PRIOR FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: EP 03016672.2
; PRIOR FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: EP 03022225.1
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 284
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-11-251-208-284

Query Match      71.1%; Score 32; DB 7; Length 514;
Best Local Similarity 85.7%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      3 YEDIAQK 9
Db      49 YEDIAQK 55

RESULT 8
US-10-953-349-4732
; Sequence 4732, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 4732
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-953-349-4732

Query Match      68.9%; Score 31; DB 6; Length 175;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY      1 AQYEDIAQK 9
Db      70 ASYENIAKK 78

RESULT 9
US-10-953-349-4731
; Sequence 4731, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
```

```
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 4731
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-953-349-4731

Query Match      68.9%; Score 31; DB 6; Length 195;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY      1 AQYEDIAQK 9
Db      91 ASYENIAKK 99

RESULT 10
US-10-566-644-36
; Sequence 36, Application US/10566644
; Publication No. US20060137042A1
; GENERAL INFORMATION:
; APPLICANT: Plesch, Gunnar
; APPLICANT: Puzio, Piotr
; APPLICANT: Blau, Astrid
; APPLICANT: Looser, Ralf
; APPLICANT: Wendel, Birgit
; APPLICANT: Kamlage, Beate
; APPLICANT: Chardonnens, Agnes
; APPLICANT: Shirley, Amber
; APPLICANT: Wang, Xi-Qing
; APPLICANT: Sarria-Millan, Rodrigo
; APPLICANT: McKersie, Bryan
; APPLICANT: Chen, Ruoying
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF FINE CHEMICALS
; FILE REFERENCE: 12810-00197-US
; CURRENT APPLICATION NUMBER: US/10/566,644
; CURRENT FILING DATE: 2006-01-31
; PRIOR APPLICATION NUMBER: EP 03016672.2
; PRIOR FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: PCT/US2004/11887
; PRIOR FILING DATE: 2004-04-15
; NUMBER OF SEQ ID NOS: 400
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 36
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Brassica napus
; US-10-566-644-36

Query Match      68.9%; Score 31; DB 6; Length 195;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY      1 AQYEDIAQK 9
Db      91 ASYENIAKK 99

RESULT 11
US-11-251-208-345
; Sequence 345, Application US/11251208
; Publication No. US20060137043A1
; GENERAL INFORMATION:
; APPLICANT: Puzio, Piotr
; APPLICANT: Chardonnens, Agnes
; APPLICANT: Shirley, Amber
; APPLICANT: Wang, Xi-Qing
; APPLICANT: Sarria-Millan, Rodrigo
```

; APPLICANT: McKersie, Bryan  
; APPLICANT: Chen, Ruoying  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING PROTEINS ASSOCIATED WITH ABIOTIC  
; TITLE OF INVENTION: STRESS RESPONSE AND PLANT CELLS AND PLANTS WITH INCREASED  
; TITLE OF INVENTION: TOLERANCE TO ENVIRONMENTAL STRESS  
; FILE REFERENCE: 13311-00015-US  
; CURRENT APPLICATION NUMBER: US/11/251,208  
; CURRENT FILING DATE: 2005-10-14  
; PRIOR APPLICATION NUMBER: PCT/US2004/011888  
; PRIOR FILING DATE: 2004-04-15  
; PRIOR APPLICATION NUMBER: EP 03008080.8  
; PRIOR FILING DATE: 2003-04-15  
; PRIOR APPLICATION NUMBER: EP 03009728.1  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: EP 03016672.2  
; PRIOR FILING DATE: 2003-08-01  
; PRIOR APPLICATION NUMBER: EP 03022225.1  
; PRIOR FILING DATE: 2003-09-30  
; NUMBER OF SEQ ID NOS: 852  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 345  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Brassica napus  
US-11-251-208-345

Query Match 68.9%; Score 31; DB 7; Length 195;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 91 ASYENIAKK 99

RESULT 12  
US-10-566-644-8  
; Sequence 8, Application US/10566644  
; Publication No. US20060137042A1  
; GENERAL INFORMATION:  
; APPLICANT: Plesch, Gunnar  
; APPLICANT: Puzio, Piotr  
; APPLICANT: Blau, Astrid  
; APPLICANT: Looser, Ralf  
; APPLICANT: Wendel, Birgit  
; APPLICANT: Kamlage, Beate  
; APPLICANT: Chardonnens, Agnes  
; APPLICANT: Shirley, Amber  
; APPLICANT: Wang, Xi-Qing  
; APPLICANT: Sarría-Millan, Rodrigo  
; APPLICANT: McKersie, Bryan  
; APPLICANT: Chen, Ruoying  
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF FINE CHEMICALS  
; FILE REFERENCE: 12810-00197-US  
; CURRENT APPLICATION NUMBER: US/10/566,644  
; CURRENT FILING DATE: 2006-01-31  
; PRIOR APPLICATION NUMBER: EP 03016672.2  
; PRIOR FILING DATE: 2003-08-01  
; PRIOR APPLICATION NUMBER: PCT/US2004/11887  
; PRIOR FILING DATE: 2004-04-15  
; NUMBER OF SEQ ID NOS: 400  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 8  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Brassica napus  
US-10-566-644-8

Query Match 68.9%; Score 31; DB 6; Length 196;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9

Db 91 ASYENIAKK 99  
RESULT 13  
US-10-566-644-16  
; Sequence 16, Application US/10566644  
; Publication No. US20060137042A1  
; GENERAL INFORMATION:  
; APPLICANT: Plesch, Gunnar  
; APPLICANT: Puzio, Piotr  
; APPLICANT: Blau, Astrid  
; APPLICANT: Looser, Ralf  
; APPLICANT: Wendel, Birgit  
; APPLICANT: Kamlage, Beate  
; APPLICANT: Chardonnens, Agnes  
; APPLICANT: Shirley, Amber  
; APPLICANT: Wang, Xi-Qing  
; APPLICANT: Sarría-Millan, Rodrigo  
; APPLICANT: McKersie, Bryan  
; APPLICANT: Chen, Ruoying  
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF FINE CHEMICALS  
; FILE REFERENCE: 12810-00197-US  
; CURRENT APPLICATION NUMBER: US/10/566,644  
; CURRENT FILING DATE: 2006-01-31  
; PRIOR APPLICATION NUMBER: EP 03016672.2  
; PRIOR FILING DATE: 2003-08-01  
; PRIOR APPLICATION NUMBER: PCT/US2004/11887  
; PRIOR FILING DATE: 2004-04-15  
; NUMBER OF SEQ ID NOS: 400  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 16  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-566-644-16

Query Match 68.9%; Score 31; DB 6; Length 196;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 91 ASYENIAKK 99

RESULT 14  
US-11-251-208-304  
; Sequence 304, Application US/11251208  
; Publication No. US20060137043A1  
; GENERAL INFORMATION:  
; APPLICANT: Puzio, Piotr  
; APPLICANT: Chardonnens, Agnes  
; APPLICANT: Shirley, Amber  
; APPLICANT: Wang, Xi-Qing  
; APPLICANT: Sarría-Millan, Rodrigo  
; APPLICANT: McKersie, Bryan  
; APPLICANT: Chen, Ruoying  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING PROTEINS ASSOCIATED WITH ABIOTIC  
; TITLE OF INVENTION: STRESS RESPONSE AND PLANT CELLS AND PLANTS WITH INCREASED  
; TITLE OF INVENTION: TOLERANCE TO ENVIRONMENTAL STRESS  
; FILE REFERENCE: 13311-00015-US  
; CURRENT APPLICATION NUMBER: US/11/251,208  
; CURRENT FILING DATE: 2005-10-14  
; PRIOR APPLICATION NUMBER: PCT/US2004/011888  
; PRIOR FILING DATE: 2004-04-15  
; PRIOR APPLICATION NUMBER: EP 03008080.8  
; PRIOR FILING DATE: 2003-04-15  
; PRIOR APPLICATION NUMBER: EP 03009728.1  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: EP 03016672.2  
; PRIOR FILING DATE: 2003-08-01  
; PRIOR APPLICATION NUMBER: EP 03022225.1

; PRIOR FILING DATE: 2003-09-30  
; NUMBER OF SEQ ID NOS: 852  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 304  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Brassica napus  
US-11-251-208-304

Query Match 68.9%; Score 31; DB 7; Length 196;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
| : : : : :  
Db 91 ASYENIAKK 99

## RESULT 15

US-11-251-208-407  
; Sequence 407, Application US/11251208  
; Publication No. US20060137043A1  
; GENERAL INFORMATION:  
; APPLICANT: Puzio, Piotr  
; APPLICANT: Chardonnens, Agnes  
; APPLICANT: Shirley, Amber  
; APPLICANT: Wang, Xi-Qing  
; APPLICANT: Sarria-Millan, Rodrigo  
; APPLICANT: Mckersie, Bryan  
; APPLICANT: Chen, Ruoying  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING PROTEINS ASSOCIATED WITH ABIOTIC  
; TITLE OF INVENTION: STRESS RESPONSE AND PLANT CELLS AND PLANTS WITH INCREASED  
; TITLE OF INVENTION: TOLERANCE TO ENVIRONMENTAL STRESS  
; FILE REFERENCE: 13311-00015-US  
; CURRENT APPLICATION NUMBER: US/11/251,208  
; CURRENT FILING DATE: 2005-10-14  
; PRIOR APPLICATION NUMBER: PCT/US2004/011888  
; PRIOR FILING DATE: 2004-04-15  
; PRIOR APPLICATION NUMBER: EP 03008080.8  
; PRIOR FILING DATE: 2003-04-15  
; PRIOR APPLICATION NUMBER: EP 03009728.1  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: EP 03016672.2  
; PRIOR FILING DATE: 2003-08-01  
; PRIOR APPLICATION NUMBER: EP 03022225.1  
; PRIOR FILING DATE: 2003-09-30  
; NUMBER OF SEQ ID NOS: 852  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 407  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-11-251-208-407

Query Match 68.9%; Score 31; DB 7; Length 196;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
| : : : : :  
Db 91 ASYENIAKK 99

Search completed: July 5, 2006, 20:57:47  
Job time : 5.125 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:24:38 ; Search time 15.9375 Seconds  
(without alignments)  
49.429 Million cell updates/sec

Title: US-10-774-928A-2

Perfect score: 45

Sequence: 1 AQVEDIAQK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/iaa/5\_COMB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/iaa/6\_COMB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/iaa/7\_COMB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/iaa/H\_COMB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/iaa/RE\_COMB.pep.\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	9	2	US-09-809-003A-2
2	45	100.0	637	2	US-09-949-016-8152
3	38	84.4	384	2	US-09-949-016-11034
4	38	84.4	564	2	US-09-949-016-6628
5	38	84.4	569	2	US-09-949-016-11035
6	38	84.4	569	2	US-09-949-016-11036
7	38	84.4	608	2	US-10-094-749-2446
8	38	84.4	645	2	US-09-919-172-41
9	37	82.2	482	2	US-09-538-092-858
10	37	82.2	483	2	US-09-919-497-79
11	37	82.2	502	2	US-09-949-016-11033
12	37	82.2	551	1	US-09-067-351-2
13	37	82.2	551	2	US-09-360-490-2
14	36	80.0	233	2	US-10-163-214-10
15	36	80.0	443	2	US-09-949-016-10582
16	36	80.0	643	2	US-09-538-092-844
17	36	80.0	914	2	US-10-163-214-12
18	35	77.8	8	2	US-09-809-003A-10
19	35	77.8	9	2	US-09-809-003A-4
20	35	77.8	315	2	US-09-543-681A-6875
21	34	75.6	398	2	US-10-104-047-3388
22	34	75.6	564	2	US-09-252-991A-24049
23	33	73.3	110	2	US-09-205-258-811
24	33	73.3	110	2	US-10-004-860-811
25	33	73.3	283	2	US-09-252-991A-21171
26	33	73.3	411	2	US-09-205-258-815

27	33	73.3	411	2	US-10-004-860-815	Sequence 815, App
28	33	73.3	471	2	US-09-949-016-10880	Sequence 10880, A
29	33	73.3	493	2	US-09-949-016-6203	Sequence 6203, Ap
30	33	73.3	528	2	US-09-949-016-9678	Sequence 9678, Ap
31	33	73.3	566	2	US-09-328-352-4809	Sequence 4809, Ap
32	33	73.3	567	2	US-10-104-047-3772	Sequence 3772, Ap
33	33	73.3	591	2	US-09-543-681A-5171	Sequence 5171, Ap
34	33	73.3	596	2	US-09-489-039A-9470	Sequence 9470, Ap
35	33	73.3	730	2	US-09-949-016-11573	Sequence 11573, A
36	33	73.3	915	2	US-10-163-214-6	Sequence 6, Appli
37	32	71.1	114	2	US-09-902-540-13061	Sequence 13061, A
38	32	71.1	232	2	US-09-919-497-78	Sequence 78, Appl
39	32	71.1	315	2	US-09-248-786A-14850	Sequence 14850, A
40	32	71.1	430	2	US-09-919-039-105	Sequence 105, App
41	32	71.1	447	2	US-09-949-016-11032	Sequence 11032, A
42	32	71.1	469	2	US-09-538-092-948	Sequence 948, App
43	32	71.1	496	2	US-09-949-016-11133	Sequence 11133, A
44	32	71.1	514	2	US-09-538-092-495	Sequence 495, App
45	32	71.1	613	2	US-09-328-352-5116	Sequence 5116, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-809-003A-2  
; Sequence 2, Application US/09809003A

; Patent No. 6673351

; GENERAL INFORMATION:

; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission

; TITLE OF INVENTION: Psoriasis

; FILE REFERENCE: 302178

; CURRENT APPLICATION NUMBER: US/09/809,003A

; CURRENT FILING DATE: 2001-03-16

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Leishmania

US-09-809-003A-2

Query Match 100.0%; Score 45; DB 2; Length 9;

Best Local Similarity 100.0%; Pred. No. 5e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDIAQK 9

Db 1 AQVEDIAQK 9

##### RESULT 2

US-09-949-016-8152

; Sequence 8152, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8152

; LENGTH: 637

```
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8152

Query Match      100.0%; Score 45; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 356 AQYEDIAQK 364

RESULT 3
US-09-949-016-11034
; Sequence 11034, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11034
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11034

Query Match      84.4%; Score 38; DB 2; Length 384;
Best Local Similarity 77.8%; Pred. No. 9.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 362 AQYEDIAQK 370

RESULT 4
US-09-949-016-6628
; Sequence 6628, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6628
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6628

Query Match      84.4%; Score 38; DB 2; Length 564;
```

```
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 339 AQYEDIAQK 347

RESULT 5
US-09-949-016-11035
; Sequence 11035, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11035
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11035

Query Match      84.4%; Score 38; DB 2; Length 569;
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 344 AQYEDIAQK 352

RESULT 6
US-09-949-016-11036
; Sequence 11036, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11036
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11036

Query Match      84.4%; Score 38; DB 2; Length 569;
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9
Db 344 AQYEDIAQK 352
```



Db 344 AQYEIAQR 352

## RESULT 7

US-10-094-749-2446  
; Sequence 2446, Application US/10094749  
; Patent No. 6979557  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KELLICH  
; APPLICANT: TRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHIKO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2446  
; LENGTH: 608  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-2446

Query Match 84.4%; Score 38; DB 2; Length 608;  
Best Local Similarity 77.8%; Pred. No. 15;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 401 AQYEIAQR 409

## RESULT 8

US-09-919-172-41  
; Sequence 41, Application US/09919172  
; Patent No. 6673545  
; GENERAL INFORMATION:  
; APPLICANT: Faris, Mary  
; APPLICANT: Turner, Christopher M.  
; TITLE OF INVENTION: PROSTATE CANCER MARKERS  
; FILE REFERENCE: PA-0036 US  
; CURRENT APPLICATION NUMBER: US/09/919,172  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 60/222,469  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PERL Program  
; SEQ ID NO 41  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673545 1798379CD1

US-09-919-172-41

Query Match 84.4%; Score 38; DB 2; Length 645;  
Best Local Similarity 77.8%; Pred. No. 16;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 360 AQYEIAQR 368

## RESULT 9

US-09-538-092-858  
; Sequence 858, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurPatSeqFormater Version 0.9  
; SEQ ID NO 858  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P05787  
US-09-538-092-858

Query Match 82.2%; Score 37; DB 2; Length 482;  
Best Local Similarity 77.8%; Pred. No. 19;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 264 AQYEDIAQR 272

## RESULT 10

US-09-919-497-79  
; Sequence 79, Application US/09919497  
; Patent No. 6773883  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 79  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-497-79

Query Match 82.2%; Score 37; DB 2; Length 483;  
Best Local Similarity 77.8%; Pred. No. 19;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 1 AQYEDIAQK 9

Db 265 AQYEDIANR 273

RESULT 11

US-09-949-016-11033

Sequence 11033, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11033

LENGTH: 502

TYPE: PRT

ORGANISM: Human

US-09-949-016-11033

Query Match 82.2%; Score 37; DB 2; Length 502;

Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAK 9

Db 284 AQYEDIANR 292

RESULT 12

US-09-067-351-2

Sequence 2, Application US/09067351

Patent No. 5994081

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Hillman, Jennifer L.

APPLICANT: Corley, Neil C.

APPLICANT: Baughn, Mariah

TITLE OF INVENTION: HUMAN KERATINS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION NUMBER: US/09/067,351

FILING DATE: Herewith

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0511 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 551 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: KERANOT02

CLONE: 2029060

US-09-360-490-2

Query Match 82.2%; Score 37; DB 2; Length 551;

Best Local Similarity 77.8%; Pred. No. 22;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAK 9

Db 325 AQYEDIANR 333

RESULT 13

US-09-360-490-2

Sequence 2, Application US/09360490

Patent No. 6221843

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Hillman, Jennifer L.

APPLICANT: Corley, Neil C.

APPLICANT: Baughn, Mariah

TITLE OF INVENTION: HUMAN KERATINS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION NUMBER: US/09/360,490

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/067,351

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0511 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 551 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: KERANOT02

CLONE: 2029060

US-09-360-490-2

Query Match 82.2%; Score 37; DB 2; Length 551;

Best Local Similarity 77.8%; Pred. No. 22;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAK 9

Db 325 AQYEDIANR 333

RESULT 14  
US-10-163-214-10  
; Sequence 10, Application US/10163214  
; Patent No. 6849781  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Broglie, Karen E.  
; APPLICANT: Butler, Karlene H.  
; APPLICANT: Thorpe, Catherine J.  
; TITLE OF INVENTION: Starch Synthase Isoform V  
; FILE REFERENCE: BE1520 US NA  
; CURRENT APPLICATION NUMBER: US/10/163,214  
; PRIOR FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: 60/297,099  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 10  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-163-214-10

Query Match 80.0%; Score 36; DB 2; Length 293;  
Best Local Similarity 77.8%; Pred. No. 18;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 276 SQYEDIYQK 284

RESULT 15  
US-09-949-016-10582  
; Sequence 10582, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10582  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10582

Query Match 80.0%; Score 36; DB 2; Length 443;  
Best Local Similarity 77.8%; Pred. No. 27;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDIAQK 9  
Db 225 AQYEDIAQK 233

Search completed: July 5, 2006, 19:26:42  
Job time : 15.9375 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:20:41 ; Search time 14.6667 Seconds  
(without alignments)  
104.964 Million cell updates/sec

Title: US-10-774-928A-3

Perfect score: 87

Sequence: 1 EIFTYHNLEGGQEDF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80:\*\*

1: Pirl:\*\*

2: Pirl:\*\*

3: Pirl:\*\*

4: Pirl:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	622	2 I37984	keratin 9, type I,
2	46	52.9	483	2 A55033	keratin 12 - mouse
3	45.5	52.3	400	1 KRH09	keratin 19, type I
4	45	51.7	399	2 A25470	cytokeatin 19 - b
5	45	51.7	419	2 A25438	keratin, type I cy
6	45	51.7	450	2 I50484	vimentin beta - go
7	45	51.7	3678	2 S28316	dystrophin - mouse
8	44.5	51.1	423	2 I59463	keratin, type I, c
9	44.5	51.1	430	2 S05481	keratin 18, type I
10	44	50.6	105	2 S55691	keratin type I (cl
11	44	50.6	368	2 A28925	keratin, type I no
12	44	50.6	728	2 A54603	transcription fact
13	44	50.6	834	2 A97178	probable permealase
14	43	49.4	135	2 PNO494	NAD ADP-ribosyltra
15	43	49.4	304	2 I50721	synenin - chicken
16	43	49.4	526	1 KRBOVI	keratin, 54K type
17	43	49.4	561	2 A31994	keratin 10, type I
18	43	49.4	569	1 KRMSE1	keratin, 59K type
19	43	49.4	570	2 S07330	keratin, epidermal
20	43	49.4	593	1 KRH00	keratin 10, type I
21	43	49.4	911	2 JC6016	chitin synthase (E
22	43	49.4	916	2 JC2315	chitin synthase (E
23	43	49.4	1011	1 JH0581	NAD ADP-ribosyltra
24	43	49.4	1028	2 I51173	myosin I beta bu
25	42	48.3	209	2 T17294	hypothetical prote
26	42	48.3	222	2 G86168	hypothetical prote
27	42	48.3	327	2 S04511	keratin 3, type I,
28	42	48.3	411	2 S45318	keratin 12 - rabbi
29	42	48.3	429	2 A25145	keratin, 47K type

30 42 48.3 436 1 A70409  
31 42 48.3 486 1 KRXL  
32 42 48.3 566 2 S42257  
33 42 48.3 566 2 T20390  
34 42 48.3 3263 2 E82410  
35 41.5 47.7 520 2 J50291  
36 41 47.1 175 2 C97138  
37 41 47.1 256 2 A95887  
38 41 47.1 257 2 I38025  
39 41 47.1 264 2 H64086  
40 41 47.1 335 2 G96607  
41 41 47.1 346 2 S34165  
42 41 47.1 394 2 B82931  
43 41 47.1 401 2 E70646  
44 41 47.1 403 2 JQ0028  
45 41 47.1 480 2 A56694

#### ALIGNMENTS

##### RESULT 1

I37984

keratin 9, type I, cytoskeletal - human

N;Alternate names: cytokeatin 9; scatter protein 60K chain, placental

C;Species: Homo sapiens (man)

C;Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text change 09-Jul-2004

C;Accession: I37984; S40307; S77921; S41161; B35494; I37943

R;Reis, A.; Hennies, H.C.; Langbein, L.; Digweed, M.; Mischke, D.; Drechsler, M.; Schroe

Nature Genet. 6, 174-179, 1994

A;Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).

A;Reference number: I37984; MUID:94214498; PMID:7512862

A;Accession: I37984

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-622 <RES>

A;Cross-references: UNIPROT:P35527; UNIPARC:UPI000012DAE1; EMBL:X75015; NID:g453154; PIDN

R;Langbein, L.; Heid, H.W.; Moll, I.; Franke, W.W.

Differentiation 55, 57-71, 1993

A;Title: Molecular characterization of the body site-specific human epidermal cytoke

A;Reference number: I37943; MUID:94131202; PMID:7507869

A;Accession: S40307

A;Molecule type: mRNA

A;Residues: 1-622 <LAN>

A;Cross-references: UNIPARC:UPI000012DAE1; EMBL:Z29074; NID:g435475; PIDN:CAA82315.1; PI

A;Accession: S77921

A;Molecule type: protein

A;Residues: 29-53;62-104;167-188;199-233;241-249;295-362;374-430;450-480;579-604 <LAF>

A;Cross-references: UNIPARC:UPI00001774A9; UNIPARC:UPI00001774AA; UNIPARC:UPI00001774AB

4B0; UNIPARC:UPI00001774B1

R;Langbein, L.

submitted to the EMBL Data Library, December 1993

A;Reference number: S41161

A;Accession: S41161

A;Molecule type: mRNA

A;Residues: 1-11,'SB',13-622 <LAW>

A;Cross-references: UNIPARC:UPI000016ABA2; EMBL:Z29074; NID:g435475; PIDN:CAA82315.1; PI

R;Rosen, E.M.; Meromsky, L.; Romero, R.; Setter, E.; Goldberg, I.

Biochem. Biophys. Res. Commun. 168, 1082-1088, 1990

A;Title: Human placenta contains an epithelial scatter protein.

A;Reference number: A35494; MUID:90267446; PMID:2140676

A;Accession: B35494

A;Molecule type: protein

A;Residues: 'X',450-465 <ROS>

A;Cross-references: UNIPARC:UPI00001774E2

C;Genetics:

A;Gene: GDB:KRT9; EPPK

A;Cross-references: GDB:303970; OMIM:144200

A;Map position: 17q12-17q21

A;Introns: 213/3; 241/2; 293/3; 347/3; 389/3; 464/2

A;Note: defects in this gene may cause epidermolytic palmoplantar keratoderma

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil; intermediate filament

F:1-153/Domain: head #status predicted <HEA>  
 F:154-459/Domain: helical rod #status predicted <ROD>  
 F:460-622/Domain: tail #status predicted <TAI>

Query Match 100.0%; Score 87; DB 2; Length 622;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQEDF 16  
 |||||  
 Db 449 EIETHNLLGGQEDF 464

RESULT 2  
 AS5033  
 keratin 12 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 18-Nov-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C:Accession: A55033  
 R:Lin, C.Y.; Zhu, G.; Converse, R.; Kao, C.W.C.; Nakamura, H.; Tseng, S.C.G.; Mui, M.M.;  
 J. Biol. Chem. 269, 24627-24636, 1994  
 A:Title: Characterization and chromosomal localization of the cornea-specific murine ker  
 A:Reference number: A55033; MUID:95014223; PMID:7523376  
 A:Accession: A55033  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-483 <LIU>  
 A:Cross-references: UNIPROT:Q64291; UNIPARC:UPI00002996F; GB:U08095; NID:G565659; PIDN:  
 A:Note: authors translated the codon ATC for residue 225 as Thr, and GCG for residue 388  
 C:Superfamily: cytoskeletal keratin

Query Match 52.9%; Score 46; DB 2; Length 483;  
 Best Local Similarity 64.3%; Pred. No. 8.2;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQEE 14  
 |||||  
 Db 418 EIETHNLLGGQDS 431

RESULT 3  
 KRH09  
 keratin 19, type I, cytoskeletal - human  
 N:Alternate names: cytokeratin 19  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Sep-1991 #sequence\_revision 23-Mar-1995 #text\_change 09-Jul-2004  
 C:Accession: A31370; A61556; A60779; S00658; S60152  
 R:Eckert, R.L.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 1114-1118, 1988  
 A:Title: Sequence of the human 40-kDa keratin reveals an unusual structure with very hig  
 A:Reference number: A31370; MUID:88124966; PMID:2448790  
 A:Accession: A31370  
 A:Molecule type: mRNA  
 A:Residues: 1-400 <ECK>  
 A:Cross-references: UNIPROT:Q96A53; UNIPARC:UPI0000071F08; GB:J03607; NID:gi84658; PIDN:  
 R:Bader, B.L.; Jahn, L.; Franke, W.W.  
 Eur. J. Cell Biol. 47, 306-319, 1988  
 A:Title: Low level expression of cytokeratins 8, 18 and 19 in vascular smooth muscle cel  
 containing the cytokeratin 19 gene.  
 A:Reference number: A61556; MUID:89210901; PMID:2468493  
 A:Accession: A61556  
 A:Molecule type: mRNA  
 A:Residues: 1-400 <BAD>  
 A:Cross-references: UNIPARC:UPI0000071F08  
 R:Stasiak, P.C.; Purkis, P.E.; Leigh, I.M.; Lane, E.B.  
 J. Invest. Dermatol. 92, 707-716, 1989  
 A:Title: Keratin 19: predicted amino acid sequence and broad tissue distribution suggest  
 A:Reference number: A60779; MUID:89235250; PMID:2469734  
 A:Accession: A60779  
 A:Molecule type: mRNA  
 A:Residues: 1-349, 'A', 351-400 <STA>  
 A:Cross-references: UNIPARC:UPI000012DAE; EMBL:Y00503; NID:g34038; PIDN:CAA68556.1; PID

Nucleic Acids Res. 15, 10058, 1987  
 A:Title: Sequence of cDNA coding for human keratin 19.  
 A:Reference number: S00658; MUID:88096504; PMID:2447559  
 A:Accession: S00658  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-75, 'H', 78-341, 'Y', 343-349, 'A', 351-400 <ST2>  
 A:Cross-references: UNIPARC:UPI0000173D4D; EMBL:Y00503  
 A:Note: this sequence has been revised  
 R:Peterson, G.; Song, D.; Huegle-Doerr, B.; Oldenburg, I.; Bautz, E.K.F.  
 Mol. Gen. Genet. 249, 425-431, 1995  
 A:Title: Mapping of linear epitopes recognized by monoclonal antibodies with gene-fragment  
 A:Reference number: S60151; MUID:96133682; PMID:8552047  
 A:Accession: S60152  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 331-349, 'A', 351-380 <PET>  
 A:Cross-references: UNIPARC:UPI0000173D4E  
 C:Comment: Keratin 19 is the smallest human keratin, as the non-alpha-helical tail preser  
 C:Genetics:  
 A:Gene: GDB:KRT19  
 A:Cross-references: GDB:120131; OMIM:148020  
 A:Map position: 17q21-17q23  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; intermediate filament  
 F:1-73/Domain: head <HEA>  
 F:74-386/Domain: helical rod #status predicted <ROD>  
 F:387-399/Domain: tail <TAI>

Query Match 52.3%; Score 45.5; DB 1; Length 400;  
 Best Local Similarity 73.3%; Pred. No. 8;  
 Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQED 15  
 |||||  
 Db 376 EIATYRSLLG-QGED 389

RESULT 4  
 A25470  
 cytokeratin 19 - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 09-Jul-2004  
 C:Accession: A25470  
 R:Bader, B.L.; Magin, T.M.; Hatzfeld, M.; Franke, W.W.  
 EMBO J. 5, 1865-1875, 1986  
 A:Title: Amino acid sequence and gene organization of cytokeratin no. 19, an exceptional  
 A:Reference number: A25470; MUID:87004553; PMID:2428612  
 A:Accession: A25470  
 A:Molecule type: DNA  
 A:Residues: 1-399 <BAD>  
 A:Cross-references: UNIPROT:P08728; UNIPARC:UPI000012DAED; GB:X04152; NID:g469; PIDN:CAA:  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil

Query Match 51.7%; Score 45; DB 2; Length 399;  
 Best Local Similarity 81.8%; Pred. No. 9.7;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIETHNLLLEG 11  
 |||||  
 Db 375 EIATYRNLLLEG 385

RESULT 5  
 A25438  
 keratin, type I cytoskeletal protein, B2 - African clawed frog (fragment)  
 C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 09-Jul-2004  
 C:Accession: A25438  
 R:Myatani, S.; Winkles, J.A.; Sargent, T.D.; Dawid, I.B.  
 J. Cell Biol. 103, 1957-1965, 1986  
 A:Title: Stage-specific keratins in Xenopus laevis embryos and tadpoles: the XK81 gene fr

A;Reference number: A25438; MUID:87057649; PMID:2430981  
A;Accession: A25438  
A;Molecule type: mRNA  
A;Residues: 1-419 <MIY>  
A;Cross-references: UNIPROT:P05781; UNIPARC:UPI000012DADF; GB:M18155; GB:X05865; NID:g21  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 51.7%; Score 45; DB 2; Length 419;  
Best Local Similarity 81.8%; Pred. No. 10;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIEYHNLLGG 11  
||||| |||||  
Db 382 EIEYRRLLEG 392

RESULT 6  
I50484  
vimentin beta - goldfish  
C;Species: Carassius auratus (goldfish)  
C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C;Accession: I50484  
R;Glasgow, E.; Druger, R.K.; Fuchs, C.; Levine, E.M.; Giordano, S.; Schechter, N.  
J. Neurochem. 63, 470-481, 1994  
A;Title: Cloning of multiple forms of goldfish vimentin: differential expression in CNS.  
A;Reference number: I50482; MUID:94308773; PMID:8035174  
A;Accession: I50484  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-450 <GLA>  
A;Cross-references: UNIPROT:P48673; UNIPARC:UPI000013881F; GB:L23841; NID:g388285; PIDN:  
C;Superfamily: cytoskeletal keratin

Query Match 51.7%; Score 45; DB 2; Length 450;  
Best Local Similarity 81.8%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIEYHNLLGG 11  
||||| |||||  
Db 382 EIAIYRNLLEG 392

RESULT 7  
S28916  
dystrophin - mouse  
N;Alternate names: duchenne muscular dystrophy protein  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Nov-1993 #sequence\_revision 03-Nov-1995 #text\_change 09-Jul-2004  
C;Accession: S28916; B27162; S10922; C43837; B40134  
R;Bies, R.D.; Phelps, S.F.; Cortez, M.D.; Roberts, R.; Caskey, C.T.; Chamberlain, J.S.  
Nucleic Acids Res. 20, 1725-1731, 1992  
A;Title: Human and murine dystrophin mRNA transcripts are differentially expressed during  
A;Reference number: S28916; MUID:92253376; PMID:1579466  
A;Accession: S28916  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-3678 <BIE>  
A;Cross-references: UNIPROT:P11531; UNIPARC:UPI00000279E7; EMBL:M68859  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
R;Koenig, M.; Hoffman, E.P.; Bertelson, C.J.; Monaco, A.P.; Feener, C.; Kunkel, L.M.  
Cell 50, 509-517, 1987  
A;Title: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary  
A;Reference number: A90897; MUID:87273512; PMID:3607877  
A;Accession: B27162  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-201 <KOE>  
A;Cross-references: UNIPARC:UPI0000177686  
R;Nudel, U.; Zuk, D.; Einat, P.; Zeelon, E.; Levy, Z.; Neuman, S.; Yaffe, D.  
Nature 337, 76-78, 1989  
A;Title: Duchenne muscular dystrophy gene product is not identical in muscle and brain.  
A;Reference number: S06461; MUID:8902658; PMID:2909892

A;Accession: S10922  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-106 <NUD>  
A;Cross-references: UNIPARC:UPI0000177687; EMBL:X14183  
R;Rapaport, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; Fe  
Differentiation 49, 187-193, 1992  
A;Title: Characterization and cell type distribution of a novel, major transcript of the  
A;Reference number: A43837; MUID:92316332; PMID:1377655  
A;Accession: C43837  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 'MREHLKG'.3069-3181 <RAP>  
A;Cross-references: UNIPARC:UPI0000177688  
R;Hoffman, E.P.; Monaco, A.P.; Feener, C.C.; Kunkel, L.M.  
Science 238, 347-350, 1987  
A;Title: Conservation of the Duchenne muscular dystrophy gene in mice and humans.  
A;Reference number: A40134; MUID:88018015; PMID:3659917  
A;Accession: B40134  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 300-676, 'F', 678-1390 <HOF>  
A;Cross-references: UNIPARC:UPI0000177689; GB:M18025  
C;Genetics: 11/1  
A;Introns: 11/1  
A;Note: the list of introns may be incomplete  
C;Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystrop  
C;Keywords: actin binding; alternative splicing  
F;14-233/Domain: alpha-actinin actin-binding domain homology <ACT>  
F;340-449/Domain: spectrin/dystrophin repeat homology <SP1>  
F;450-558/Domain: spectrin/dystrophin repeat homology <SP2>  
F;2797-2924/Domain: spectrin/dystrophin repeat homology <SP3>  
F;3048-3085/Domain: WW repeat homology <WW1>

Query Match 51.7%; Score 45; DB 2; Length 3678;  
Best Local Similarity 61.1%; Pred. No. 13e+02;  
Matches 11; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 EIEI-----YHNLEGGQEE 14  
||||| ||||| |||||  
Db 2737 EIEIHTDIYHNLDENGQK 2754

RESULT 8  
I59463  
keratin, type I, cytoskeletal - mouse  
N;Alternate names: endo B cyokeratin; keratin D  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 09-Jul-2004  
C;Accession: I59463; A25621; A28428; JT0406  
R;Alonso, A.; Weber, T.; Jorcano, J.L.  
Roux's Arch. Dev. Biol. 196, 16-21, 1987  
A;Title: Cloning and characterization of keratin D, a murine endodermal cytoskeletal prot  
A;Reference number: I59463  
A;Accession: I59463  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-423 <RES>  
A;Cross-references: UNIPROT:P05784; UNIPARC:UPI000016CE78; GB:M36376; NID:g198587; PIDN:  
R;Singer, P.A.; Trevor, K.; Oshima, R.G.  
J. Biol. Chem. 261, 538-547, 1986  
A;Title: Molecular cloning and characterization of the endo B cyokeratin expressed in p  
A;Reference number: A25621; MUID:86085876; PMID:2416755  
A;Accession: A25621  
A;Molecule type: mRNA  
A;Residues: 1-243, 'D', 245-252, 'A', 254-423 <SIN>  
A;Cross-references: UNIPARC:UPI0000164A33; GB:M1686; NID:g198620; PIDN:AAA39390.1; PID:  
R;Oshima, R.G.; Trevor, K.; Shevinsky, L.H.; Ryder, O.A.; Cecena, G.  
Genes Dev. 2, 505-516, 1988  
A;Title: Identification of the gene coding for the endo B murine cyokeratin and its met  
A;Reference number: A28428; MUID:88255838; PMID:2454868  
A;Accession: A28428

```

A:Molecule type: DNA
A:Residues: 1-132 <OSH>
A:Cross-references: UNIPARC:UPI000011E072; GB:Y00217; NID:G50842; PIDN:CAA68365.1; PID:G50842
R:Chinese, Y.; Morita, T.; Zhang, F.; Srimahasingram, S.; Tondella, M.L.C.; Matsumoto,
Gene 70, 85-95, 1998
A:Title: Nucleotide sequence and structure of the mouse cytokelekin endoB gene.
A:Reference number: J70406; MUID:89196920; PMID:2467843
A:Accession: J70406
A:Molecule type: DNA
A:Residues: 1-133, 'F', 135-243, 'D', 245-252, 'A', 254-423 <ICH>
A:Cross-references: UNIPARC:UPI000016CD33; GB:M22832; NID:G340757; PIDN:AAA37552.1; PID:
C:Genetics:
A:Gene: endoB; KERG
A:Introns: 132/3; 160/2; 212/3; 267/3; 309/3; 384/2
C:Superfamily: cytokelekin keratin
C:Keywords: coiled coil; intermediate filament

Query Match 51.1%; Score 44.5; DB 2; Length 423;
Best Local Similarity 68.8%; Pred. No. 13;
Matches 11; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQEDF 16
||| ||| ||| ||| |||
DB 369 EIATYRRLLEDG-EDF 383

RESULT 9
S05481
keratin 18, type I, cytoskeletal - human
N:Alternate names: cytokelekin 18
C:Species: Homo sapiens (man)
C:Date: 07-Sep-1990 #sequence revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: S05481; S05482; S06889
R:Oshima, R.G.; Millan, J.L.; Cecena, G.
Differentiation 33, 61-68, 1986
A:Title: Comparison of mouse and human keratin 18: a component of intermediate filaments
A:Reference number: S05481; MUID:87134778; PMID:2434380
A:Accession: S05481
A:Molecule type: mRNA
A:Residues: 1-430 <OSH>
A:Cross-references: UNIPROT:P05783; UNIPARC:UPI000004284B; EMBL:X12881; NID:G34036; PIDN:
R:Romano, V.; Hatzfeld, M.; Magin, T.M.; Zimbelmann, R.; Franke, W.W.; Maier, G.; Ponst
Differentiation 30, 244-253, 1986
A:Title: Cytokelekin expression in simple epithelia. I. Identification of mRNA coding fo
A:Reference number: S05482; MUID:86193258; PMID:2422083
A:Accession: S05482
A:Molecule type: mRNA
A:Residues: 199-201, 'Q', 203-245, 'S', 247-308, 'R', 310-311, 'R', 313-430 <ROM>
A:Cross-references: UNIPARC:UPI000016AB98; EMBL:X12876; NID:G34034; PIDN:CAA31369.1; PID
A:Note: part of this sequence was confirmed by protein sequencing
R:Leube, R.E.; Bosch, F.X.; Romano, V.; Zimbelmann, R.; Hoefler, H.; Franke, W.W.
Differentiation 33, 69-85, 1986
A:Title: Cytokelekin expression in simple epithelia.
A:Reference number: S06888; MUID:87134779; PMID:2434381
A:Accession: S06888
A:Molecule type: mRNA
A:Residues: 7-430 <LEU>
A:Cross-references: UNIPARC:UPI000016A768; EMBL:X12883; NID:G30310; PIDN:CAA31377.1; PID
C:Genetics:
A:Gene: GDB:KRT18
A:Cross-references: GDB:120127; OMIM:148070
A:Map position: 17p12-17p11
C:Superfamily: cytokelekin keratin
C:Keywords: coiled coil; intermediate filament
F:2/430/Product: keratin 18, type I, cytoskeletal #status predicted <MAT>

Query Match 51.1%; Score 44.5; DB 2; Length 430;
Best Local Similarity 68.8%; Pred. No. 13;
Matches 11; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQEDF 16
||| ||| ||| ||| |||
DB 376 EIATYRRLLEDG-EDF 390

```

## RESULT 10

S55691

keratin type I (clone Jn7-4) - African clawed frog (fragment)

C:Species: Xenopus laevis (African clawed frog)

C:Date: 28-Oct-1995 #sequence\_revision 07-Feb-1997 #text\_change 09-Jul-2004

C:Accession: S55691

R:Adati, N.; Ito, T.; Koga, C.; Kito, K.; Sakaki, Y.; Shiokawa, K.

Biochim. Biophys. Acta 1262, 43-51, 1995

A:Title: Differential display analysis of gene expression in developing embryos of Xenopus

A:Reference number: S55691; MUID:95290493; PMID:7772598

A:Accession: S55691

A:Molecule type: mRNA

A:Residues: 1-105 &lt;ADA&gt;

A:Cross-references: UNIPROT:Q91404; UNIPARC:UPI00000FC2C7; EMBL:S78089; NID:G9992222; PIDN:

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

Query Match 50.6%; Score 44; DB 2; Length 105;

Best Local Similarity 60.0%; Pred. No. 3;

Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQED 15

||| ||| ||| ||| |||

DB 51 EIQTYRRLLEDGSPD 65

## RESULT 11

A28825

keratin, type I nonepidermal - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 09-Jul-2004

C:Accession: A28825

R:Laflamme, S.E.; Jamrich, M.; Richter, K.; Sargent, T.D.; Dawid, I.B.

Genes Dev. 2, 853-862, 1988

A:Title: Xenopus endo B is a keratin preferentially expressed in the embryonic notochord.

A:Reference number: A28825; MUID:89092007; PMID:2463213

A:Accession: A28825

A:Molecule type: mRNA

A:Residues: 1-368 &lt;LAF&gt;

A:Cross-references: UNIPROT:P08802; UNIPARC:UPI000012DAEC; GB:Y00230; NID:G64863; PIDN:CI

C:Genetics:

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 50.6%; Score 44; DB 2; Length 368;

Best Local Similarity 60.0%; Pred. No. 13;

Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQED 15

||| ||| ||| ||| |||

DB 314 EIQTYRRLLEDGSPD 328

## RESULT 12

A54603

transcription factor CTCF - chicken

N:Alternate names: finger protein CTCF

C:Species: Gallus gallus (chicken)

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004

C:Accession: A54603; S35919

R:Klenova, E.M.; Nicolas, R.H.; Paterson, H.F.; Carne, A.F.; Heath, C.M.; Goodwin, G.H.;

Mol. Cell. Biol. 13, 7612-7624, 1993

A:Title: CTCF, a conserved nuclear factor required for optimal transcriptional activity

A:Reference number: A54603; MUID:94067122; PMID:8246978

A:Accession: A54603

A:Molecule type: mRNA

A:Residues: 1-728 &lt;KL6&gt;

A:Cross-references: UNIPROT:C08705; UNIPARC:UPI0000128608; EMBL:Z22605; NID:G396093; PIDN:

C:Comment: This transcription factor with eleven zinc finger motifs binds specifically to

C:Comment: The detection of multiple isoforms by Western blotting and of multiple mRNA sp



C;Keywords: alternative splicing; DNA binding; nucleus; transcription factor

Query Match 50.6%; Score 44; DB 2; Length 728;  
Best Local Similarity 60.0%; Pred. No. 29;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQED 15  
| : ||| |||||  
Db 21 ERKYORRRGGQED 35

## RESULT 13

A97178  
Probable permease [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: A97178  
R;Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: A97178  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-834 <KUR>  
A;Cross-references: UNIPROT:Q97GM0; UNIPARC:UPI000000CA47D; GB:AF001437; PIDN:AAK80212.1;  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC2255

Query Match 50.6%; Score 44; DB 2; Length 834;  
Best Local Similarity 66.7%; Pred. No. 34;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 YHNLLGGQEDF 16  
| ||||| |||||  
Db 47 YRNLLGSGYEDF 58

## RESULT 14

PN0494  
NAD ADP-ribosyltransferase (EC 2.4.2.30) - cherry salmon (fragment)  
N;Alternate names: poly ADP-ribose polymerase  
C;Species: Oncorhynchus masou (cherry salmon)  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 31-Dec-2004  
C;Accession: PN0494  
R;Ozawa, Y.; Uchida, K.; Uchida, M.; Ami, Y.; Kushida, S.; Okada, N.; Miwa, M.  
Biochem. Biophys. Res. Commun. 193, 119-125, 1993  
A;Title: Isolation of cDNAs encoding the catalytic domain of poly(ADP-ribose)polymerase  
A;Reference number: PN0494; MUID:93277538; PMID:8503897  
A;Accession: PN0494  
A;Molecule type: mRNA  
A;Residues: 1-135 <OZA>  
A;Cross-references: UNIPROT:Q08824; UNIPARC:UPI00001320E4  
C;Comment: This zinc-finger protein plays a role in DNA repair, cell growth, and differentiation  
C;Superfamily: poly(ADP-ribose) polymerase  
C;Keywords: DNA binding; glycosyltransferase; hexosyltransferase; NAD; nucleus; pentosyl

Query Match 49.4%; Score 43; DB 2; Length 135;  
Best Local Similarity 53.3%; Pred. No. 6;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQED 15  
| : ||| |||||  
Db 12 DIEVAYSLKGAED 26

## RESULT 15

150721  
synemin - chicken (fragment)  
C;Species: Gallus gallus (chicken)  
C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 31-Dec-2004

C;Accession: I50721  
R;Becker, B.; Bellin, R.M.; Sernett, S.W.; Huiatt, T.W.; Robson, R.M.  
Biochem. Biophys. Res. Commun. 213, 796-802, 1995  
A;Title: Synemin contains the rod domain of intermediate filaments.  
A;Reference number: I50721; MUID:95382823; PMID:7654240  
A;Accession: I50721  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-304 <BEC>  
A;Cross-references: UNIPROT:Q90662; UNIPARC:UPI00001774FF; EMBL:U28143; MUID:91052880; PI

Query Match 49.4%; Score 43; DB 2; Length 304;  
Best Local Similarity 81.8%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIETHNLLLEG 11  
| ||||| |||||  
Db 289 EIETVRALLEG 299

Search completed: July 5, 2006, 19:25:12  
Job time : 15.6667 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:01:48 ; Search time 186 Seconds  
(without alignments)  
79.571 Million cell updates/sec

Title: US-10-774-928A-3  
Perfect score: 87  
Sequence: 1 EIEYHNLEGGQEDF 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	623	1 K1C9 HUMAN	P35527 homo sapien
2	70	80.5	618	2 Q8CIS9 RAT	Q8CIS9 rattus norv
3	70	80.5	743	2 Q6RHW0 MOUSE	Q6RHW0 mus musculu
4	70	80.5	786	2 O18740 CANFA	O18740 canis famil
5	53	60.9	384	2 Q6DBU4 BRARE	Q6DBU4 brachydanio
6	51	58.6	431	2 Q566F9 XENLA	Q566F9 xenopus lae
7	51	58.6	431	2 Q6GLQ7 XENLA	Q6GLQ7 xenopus lae
8	51	58.6	628	2 Q6FW44 CANGA	Q6FW44 candida gla
9	50	57.5	439	2 Q9BLX8 CIOIN	Q9BLX8 ciona intes
10	49	56.3	760	2 Q65KY8 BACLD	Q65KY8 bacillus li
11	48	55.2	453	1 K1C15 SHEEP	O77727 ovis aries
12	47	54.0	829	2 Q54DN9 DICDI	Q54DN9 dictyosteli
13	47	54.0	932	2 Q8ILF1 PLAF7	Q8ILF1 plasmodium
14	46	52.9	170	2 Q9S954 MESCR	Q9S954 mesembryant
15	46	52.9	447	1 K1C15 RAT	Q64291 mus musculu
16	46	52.9	483	1 K1C12 MOUSE	Q8GUB5 mesembryant
17	46	52.9	487	2 Q8AMB2 LAMFL	Q8AMB2 lactobacill
18	46	52.9	524	2 Q5RDU9 LACAC	Q5RDU9 trypanosoma
19	46	52.9	535	2 Q4DVU3 TRYCR	Q4DVU3 plasmodium
20	46	52.9	1009	2 Q8IIIV PLAF7	Q8IIIV bacillus li
21	46	52.9	1081	2 Q65KN1 BACLD	Q65KN1 homo sapien
22	46	52.9	1172	2 Q65KN1 BACLD	P08727 homo sapien
23	45.5	52.3	400	1 K1C19 HUMAN	Q5R859 pongo pygma
24	45.5	52.3	400	1 K1C19 PONPY	Q331M1 methanospir
25	45	51.7	131	2 Q33LM1 METHU	Q33LM1 mus musculu
26	45	51.7	221	2 Q3TBO1 MOUSE	Q3TBO1 mus musculu
27	45	51.7	360	2 Q3UF47 MOUSE	Q3UF47 campylobact
28	45	51.7	386	2 Q4HM53 CAMLA	P08728 bos taurus
29	45	51.7	399	1 K1C19 BOVIN	P05781 xenopus lae
30	45	51.7	419	1 K1C4 XENLA	Q5PR28 brachydanio
31	45	51.7	438	2 Q5PR28 BRARE	

32	45	51.7	450	1 VIMB CARAU	P48673 carassius a
33	45	51.7	452	1 K1C15 MOUSE	Q61414 mus musculu
34	45	51.7	452	2 Q6IMF1 RAT	Q6IMF1 rattus norv
35	45	51.7	454	2 Q4QR57 XENLA	Q4QR57 xenopus lae
36	45	51.7	456	2 Q6IFW5 RAT	Q6IFW5 rattus norv
37	45	51.7	469	2 Q7ZV83 XENLA	Q7ZV83 xenopus lae
38	45	51.7	488	2 Q5M8X9 XENTR	Q5M8X9 xenopus tro
39	45	51.7	491	2 Q8AV12 XENLA	Q8AV12 xenopus lae
40	45	51.7	567	2 Q5MIV0 AEDAL	Q5MIV0 aedes albop
41	45	51.7	582	2 Q3NXA5 9GAMM	Q3NXA5 shewanella
42	45	51.7	587	2 Q3P3X2 9GAMM	Q3P3X2 shewanella
43	45	51.7	605	2 Q4UG78 THEAN	Q4UG78 theileria a
44	45	51.7	633	2 Q4N919 THEPA	Q4N919 theileria p
45	45	51.7	683	2 Q4RQR9 TETNG	Q4RQR9 tetraodon n

## ALIGNMENTS

RESULT 1  
K1C9 HUMAN  
ID K1C9 HUMAN STANDARD; PRT; 623 AA.  
AC P35527; O00109; Q14665;  
DT 01-JUN-1994, integrated into UniProtKB/Swiss-Prot.  
DT 08-NOV-2005, sequence version 2.  
DT 07-FEB-2006, entry version 50.  
DE Keratin, type I cytoskeletal 9 (Cytokeratin-9) (CK-9) (Keratin-9)  
DE (K9).  
GN Name=KRT9;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND PARTIAL PROTEIN SEQUENCE.  
RC TISSUE=Foot sole tissue;  
RX MEDLINE=94131202; PubMed=7507869;  
RA Langbein L., Heid H.W., Moll I., Franke W.W.;  
RT "Molecular characterization of the body site-specific human epidermal  
cytokeratin 9: cDNA cloning, amino acid sequence, and tissue  
specificity of gene expression.";  
RL Differentiation 55:57-72(1993).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANTS EPPK LYS-161; GLN-163  
AND TRP-163.  
RX MEDLINE=94214498; PubMed=7512862;  
RA Reis A., Hennies H.-C., Langbein L., Digweed M., Mischke D.,  
Dreschler M., Schroek E., Royer-Pokora B., Franke W.W., Sperling K.,  
Kuester W.;  
RT "Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma  
(EPPK).";  
RL Nat. Genet. 6:174-179(1994).  
RN [3]  
RP PROTEIN SEQUENCE OF 14-29, AND MASS SPECTROMETRY.  
RC TISSUE=Cervix carcinoma;  
RA Bienvenut W.V.;  
RL Unpublished observations (AUG-2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [MRNA] OF 147-372, AND VARIANT EPPK GLN-163.  
RX MEDLINE=96228052; PubMed=8647270; DOI=10.1016/0014-5793(96)00393-6;  
RA Kobayashi S., Tanaka T., Matsuyoshi N., Imamura S.;  
RT "Keratin 9 point mutation in the pedigree of epidermolytic hereditary  
palmoplantar keratoderma perturbs keratin intermediate filament  
network formation.";  
RL FEBS Lett. 386:149-155(1996).  
RN [5]  
RP PROTEIN SEQUENCE OF 450-466.  
RX MEDLINE=90267446; PubMed=2140676;  
RA Rosen E.M., Meromsky L., Romero R., Setter E., Goldberg I.;  
RT "Human placenta contains an epithelial scatter protein.";  
RL Biochem. Biophys. Res. Commun. 168:1082-1088(1990).  
RN [6]



```

DR PRINTS; PRO1248; TYPE1KERATIN.
KW Keratin.
SQ SEQUENCE 743 AA; 72516 MW; 76405B82F1C750B0 CRC64;

Query Match 80.5%; Score 70; DB 2; Length 743;
Best Local Similarity 81.2%; Pred. No. 0.018;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY ` 1 EIETHNLLGGQEDF 16
      ||||| |||||:|
      428 EIETYRELGGQDF 443

RESULT 4
OI18740 CANFA
ID ID OI18740 CANFA PRELIMINARY; PRT; 786 AA.
AC O18740;
DT 01-JAN-1998, integrated into UniprotKB/TREMBL.
DT 01-JAN-1998, sequence version 1.
DT 07-FEB-2006, entry version 25.
DE Keratin.
DE DE
GN Name=kRT9;
OS Canis familiaris (dog) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OC NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RP Lachume P., Htte C., Jouquand S., Priat C., Galibert F.;
RT "Identification and analysis of the dog keratin 9 (KRT9) gene.";
RL Idem. Genet. 9:173-178(1998) .

```

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
Distributed under the Creative Commons Attribution-NoDerivs License

DR HSSP; P08670; IGK7. Canis familiaris.  
DR Ensembl; ENSCAF00000015952; Canis familiaris.  
DR GO; GO:0005198; F: structural molecule activity; IEA.  
DR

```

DR interPro; IRO002957; Keratin_I.  

DR Pfam; PF00038; Filament; 1.  

DR PRINTS; PR01228; EGGSHELL.  

DR PRINTS; PR01248; TYPEIKERATIN.  

SO SEQUENCE 786 AA; 76354 MW; 272AB5425DD09535 CRC64;  

  

Query Match      80.5%; Score 70; DB 2; Length 786;  

Best Local Similarity 81.2%; Pred. No. 0.019;  

Matches 13; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  

  

QY      1 EIETHYNLLEGGQEDF 16  

       ||| :|||||||  

Db     434 EIKTYSLLGGQEDF 449  

  

RESULT 5  

Q6DBU4 BRARE  

ID ID Q6DBU4 BRARE PRELIMINARY; PRT; 384 AA.  

AC AC Q6DBU4;  

DT 16-AUG-2004, integrated into UniprotKB/TREMBL.  

DT 16-AUG-2004, sequence version 1.  

DT 07-FEB-2006, entry version 15.  

DE Zgc:92380.  

GN ORFNAMES=zgc:92380;  

OS Brachydanio rerio (zebrafish) (Danio rerio).  

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  

OC Cyprinidae; Danio.  

NCBI_TaxId=7955;  

[1]  

RP NUCLEOTIDE SEQUENCE.
```

```

RC TISSUE=Whole;
RX MEDLINE=223388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulláhy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettanan M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska J., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Whole;
RA Director MGC Project;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; BC078359, AAH78359.1; -: mRNA.
DR Ensembl; ENSDARG0000041339; Danio rerio.
DR ZFIN; ZDB-GENS-040801-221; zgc:92380.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; UNKNOWN_1.
KW Intermediate filament.
SQ SEQUENCE 384 AA; 43462 MW; 2DC33996DAE256657 CRC64;

Query Match 60.9%; Score 53; DB 2; Length 384;
Best Local Similarity 66.7%; Pred. No. 5.7;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps

QY 1 EITYHNLEGGQED 15
Db 354 EIDTYHRLLEGVDD 368
||:|||||:|
||:|||||:|

RESULT 6
ID Q566F9_XENLA PRELIMINARY; PRT; 431 AA.
AC Q566F9;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE MGC84388 protein.
OS Xenopus laevis (African clawed frog).
GN Names=MGC84388;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,

```

RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative";  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [2]  
 RN  
 RP NUCLEOTIDE SEQUENCE.  
 RC  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Dlatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Sapletchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Whiting J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Pating M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RN  
 RP NUCLEOTIDE SEQUENCE.  
 RC  
 RC TISSUE=Brain;  
 RL Klein S., Gerhard D.S.;  
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 CC EMBL; BC093567; AAH93567.1; -; mRNA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin\_I.  
 DR Pfam; PF00038; Filament; I.  
 DR PRINTS; PR01248; TYPEKERATIN.  
 DR PROSITE; PS00226; IF; UNKNOWN 1.  
 SQ SEQUENCE 431 AA; 47704 MW; 13EBE9019731326 CRC64;  
  
 Query Match 58.6%; Score 51; DB 2; Length 431;  
 Best Local Similarity 76.9%; Pred No. 14;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps  
  
 Qy 1 EIETHNLEGGQ 13  
 || ||| |||||  
 Db 381 EISTYHKLGGQQ 393  
  
 RESULT 7  
 Q6GLQ7\_XENLA PRELIMINARY; PRT; 431 AA.  
 ID Q6GLQ7\_XENLA  
 AC Q6GLQ7;  
 DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 15.  
 DR MGC84388 protein.  
 GN Name=MGC84388;  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RN  
 RP NUCLEOTIDE SEQUENCE.  
 RC  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RN [2]

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abranson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J.J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J.J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S., Krzyzinski M.I., Skalska U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2].  
 RN NUCLEOTIDE SEQUENCE.  
 RP TISSUE=Brain;  
 RC MEDLINE=22341132; PubMed=12454917; DOI=10.1002/gvdy.10174;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative.";  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [3].  
 RN NUCLEOTIDE SEQUENCE.  
 RP TISSUE=Brain;  
 RA Klein S., Gerhard D.S.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; BC074402; AAH74402.1; -; mRNA.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin\_1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPEKERATIN.  
 DR PROSITE; PS00226; IF; UNKNOWN\_1.  
 DR Intermediate filament.  
 SK SEQUENCE 431 AA; 47723 MW; 0032E596D1B243C8 CRC64;  
 Query Match 58.6%; Score 51; DB 2; Length 431;  
 Best Local Similarity 76.9%; Pred. No. 14;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 EIETYNLLEGGQ 13  
 Db 381 EISTYHKLLEGGQ 393  
 RESULT 8  
 Q6FW44\_CANGA PRELIMINARY; PRT; 528 AA.  
 ID Q6FW44\_CANGA  
 AC Q6FW44;  
 DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 11.  
 DE Candida glabrata strain CBS138 chromosome D complete sequence.  
 GN OrderedLocusNames=CAGL0D03080g;  
 OS Candida glabrata (yeast) (Torulopsis glabrata).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5478;  
 RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=ATCC 2001 / CBS 138;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,  
 RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 RA Boismare A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Poitier S., Richard G.-F., Straub M.-L., Suteau A.,  
 RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zengou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.-L.;  
 RT "Genome evolution in yeasts.";  
 RL Nature 430:35-44(2004).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; CR380950; CAG58461.1; -; Genomic\_DNA.  
 DR Complete proteome.  
 KW SEQUENCE 628 AA; 73024 MW; 5B61FC1B3D9AB450 CRC64;  
 Query Match 58.6%; Score 51; DB 2; Length 628;  
 Best Local Similarity 60.0%; Pred. No. 22;  
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 IETYNLLEGGQEDF 16  
 Db 436 IELYHNLLQGLKDY 450  
 RESULT 9  
 Q9BLX8\_CIOIN PRELIMINARY; PRT; 439 AA.  
 ID Q9BLX8\_CIOIN  
 AC Q9BLX8;  
 DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 19.  
 DE Intermediate filament protein IF-C.  
 GN Name=if-c;  
 OS Ciona intestinalis.  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 OC Phlebobranchia; Clonidae; Ciona.  
 OX NCBI\_TaxID=7719;  
 RN [1].  
 RN NUCLEOTIDE SEQUENCE.  
 RP TISSUE=Whole animal;  
 RC MEDLINE=22061815; PubMed=12067066;  
 RA Wang J., Karabinos A., Zimek A., Meyer M., Riemer D., Hudson C.,  
 RA Lemaire P., Weber K.;  
 RT "Cytoplasmic intermediate filament protein expression in tunicate  
 RT development; a specific marker for the test cells.";  
 RL Eur. J. Cell Biol. 81:302-311(2002).  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AJ298331; CAC24552.1; -; mRNA.  
 DR HSSP; P08670; 1GK7.  
 DR Ensembl; ENSCING0000008941; Ciona intestinalis.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin\_1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPEKERATIN.  
 DR PROSITE; PS00226; IF; UNKNOWN\_1.

KW Intermediate filament.  
SQ SEQUENCE 439 AA; 49875 MW; 22F60922B6D2E9D1 CRC64;  
Query Match 57.5%; Score 50; DB 2; Length 439;  
Best Local Similarity 68.8%; Pred. No. 21;  
Matches 11; Conservative 1; Mismatches 2; Indels 2; Gaps 1;  
QY 1 EIETHYHNLEGGQEDF 16  
DB 385 EISTYHKLEG--EEF 398  
RESULT 10  
Q65KY8 BACLD PRELIMINARY; PRT; 760 AA.  
AC Q65KY8\_BACLD Q62WDS;  
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.  
DT 25-OCT-2004, sequence version 1.  
DT 07-FEB-2006, entry version 15.  
DE YesS (Probable transcriptional regulator).  
GN Name=YesS; OrderedLocustNames=BL03790, BLi01372;  
OS Bacillus licheniformis (strain DSM 13 / ATCC 14580).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=279010;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed15393718; DOI=10.1159/000079829;  
RA Veith B., Herzberg C., Steckel S., Feesche J., Maurer K.H.,  
RA Ehrenreich P., Baumer S., Henne A., Liesegang H., Merkl R.,  
RA Ehrenreich A., Gottschalk G.  
RT "The complete genome sequence of Bacillus licheniformis DSM13, an  
organism with great industrial potential."  
RL J. Mol. Microbiol. Biotechnol. 7:204-211(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed15461803; DOI=10.1186/gb-2004-5-10-r77;  
RA Rey M.W., Ramaiya P., Nelson B.A., Brody-Karpin S.D., Zaretzky E.J.,  
RA Tang M., Lopez de Leon A., Xiang H., Gusti V., Clausen I.G.,  
RA Olsen P.B., Rasmussen M.D., Andersen J.T., Joergensen P.L.,  
RA Larsen T.S., Sorokin A., Bolotin A., Lapidus A., Galleron N.,  
RA Ehrlich S.D., Berka R.M.;  
RT "Complete genome sequence of the industrial bacterium Bacillus  
licheniformis and comparisons with closely related Bacillus species.";  
RL Genome Biol. 5:RESEARCH077.1-RESEARCH077.12(2004).  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
DR EMBL; AE017333; AAU40276.1; -; Genomic DNA.  
DR EMBL; CP000002; AAU22923.1; -; Genomic DNA.  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR012287; Homeodomain-rel.  
DR InterPro; IPR000005; HTHArac.  
DR Pfam; PF00165; HTH\_Arac; 2.  
DR SMART; SM00342; HTH\_ARAC; 1.  
DR PROSITE; PS00041; HTH\_ARAC\_FAMILY\_1; UNKNOWN 1.  
DR PROSITE; PS01124; HTH\_ARAC\_FAMILY\_2; 1.  
KW Complete proteome; DNA-binding; Transcription;  
KW Transcription regulation.  
SQ SEQUENCE 760 AA; 87513 MW; 4E8D8B27EAC7BDA CRC64;

Query Match 56.3%; Score 49; DB 2; Length 760;  
Best Local Similarity 66.7%; Pred. No. 58;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 2 EIETHYHNLEGGQ 13  
DB 147 LKAYHSLLEGGQ 158  
:::|||||  
:::|||||

RESULT 11

K1C15 SHEEP  
ID K1C15\_SHEEP STANDARD; PRT; 453 AA.  
AC Q77727;  
DT 06-DEC-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-NOV-1998, sequence version 1.  
DT 07-FEB-2006, entry version 25.  
DE Keratin, type I cytoskeletal 15 (Cytokeratin-15) (CK-15) (Keratin-15) (K15).  
DE K15.  
GN Name=KRT15; (Sheep).  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
OC Pecora; Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND TISSUE SPECIFICITY.  
RC TISSUE=Splice;  
RX MEDLINE=99026083; PubMed=9806795; DOI=10.1006/excr.1998.4217;  
RA Whitbread L.A., Powell B.C.;  
RT "Expression of the intermediate filament keratin gene, K15, in the  
basal cell layers of epithelia and the hair follicle.";  
RL Exp. Cell Res. 244:448-459(1998).  
CC -!- SUBUNIT: Heterotetramer of two type I and two type II keratins.  
CC -!- TISSUE SPECIFICITY: Expressed in the basal cell layers of several  
stratified epithelia including oesophagus, tongue, stomach,  
epidermis and hair follicle. In the hair follicle, expression is  
detected mainly in the basal layer of the outer root sheath (ORS),  
except just above the follicle bulb where it occurs throughout  
its thickness. Low expression levels are seen in the single layer  
of ORS cells around the base of the follicle which increases in  
the palisade-like cells of the bulb. Also expressed in the basal  
cells of the sebaceous glands, and expression in the epidermis  
occurs in a punctate pattern.  
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
microfibrillar keratin, I (acidic; 40-55 kDa) [K9 to K20] and II  
(neutral to basic 56-70 kDa) [K1 to K8].  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
DR EMBL; AJ006277; CAA06944.1; -; Genomic DNA.  
DR HSSP; P08670; 1GK7.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Coiled coil; Intermediate filament; Keratin.  
FT CHAIN 1 453  
FT Keratin, type I cytoskeletal 15.  
FT /FTId=PRO\_0000063660.  
FT REGION 1 102  
FT REGION 103 411  
FT REGION 103 138  
FT REGION 139 157  
FT REGION 158 249  
FT REGION 250 269  
FT REGION 270 411  
FT REGION 412 453  
FT REGION 412 101  
FT COMPIAS 14 101  
FT COMPIAS 426 431  
FT Poly-Gly.  
SQ SEQUENCE 453 AA; 48770 MW; E90D5239DA45015A CRC64;  
Query Match 55.2%; Score 48; DB 1; Length 453;  
Best Local Similarity 81.8%; Pred. No. 47;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EIETHYHNLEGG 11  
DB 400 EIATYHSLLEG 410  
|||  
|||

RESULT 12



01-MAR-2003, sequence version 1.  
07-FEB-2006, entry version 10.  
Hypothetical protein.  
ORFNames=PF14\_0293;  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NCBI\_TaxID=36329;  
[1]  
NUCLEOTIDE SEQUENCE.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hymn R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Partea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McPadden G.T., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.G.;  
RT "Genome sequence of the human malaria parasite Plasmodium  
RT falciparum";  
RL Nature 419:498-511(2002).  
-----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC ENBL; AEO14820; AAN36906.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 992 AA; 116331 MW; 518F919097EFCF09 CRC64;  
  
Query Match 54.0%; Score 47; DB 2; Length 992;  
Best Local Similarity 60.0%; Pred. No. 1.7e+02;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 2 IETYNLLEGGQDF 16  
:|||||  
Db 142 LDTYENLYEGILDF 156  
  
RESULT 14  
Q9S954 MESCRC PRELIMINARY; PRT; 170 AA.  
ID Q9S954 MESCRC  
AC Q9S954;  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 01-MAY-2000, sequence version 1.  
DT 07-FEB-2006, entry version 18.  
DE Vacuolar H(+) ATPase subunit B (Fragment).  
OS Mesembryanthemum crystallinum (Common ice plant).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
OC Caryophyllales; Alzooceae; Mesembryanthemum.  
OC NCBI\_TaxID=3544;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=96151503; PubMed=8587987; DOI=10.1104/pp.110.1.1259;  
RA Low R., Rockel B., Kirsch M., Ratajczak R., Hortensteiner S.,  
RA Martinoia E., Lutge U., Rausch T.,  
RT "Early salt stress effects on the differential expression of vacuolar  
RT H(+)-ATPase genes in roots and leaves of Mesembryanthemum  
RT crystallinum";  
RT Plant Physiol. 110:259-265(1996).  
-----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.  
DR GO; GO:0005242; F:ATP binding; IEA.  
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; IEA.  
DR GO; GO:0046961; F:hydrogen-transporting ATP synthase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.  
DR InterPro: IPR000194; ATPase a/bcentre.

DR Pfam; PF00006; ATP-synt\_ab; 1.

KW Hydrolase.

FT NON\_TER 1 170

FT NON\_TER 170 170

SQ SEQUENCE 170 AA; 19062 MW; 91AC0AG3526E4BA2 CRC64;

Query Match 52.9%; Score 46; DB 2; Length 170;

Best Local Similarity 57.1%; Pred. No. 33;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 IETVHNLEGGQED 15

Db 52 LKTEENLMGGQED 65

RESULT 15

KIC15\_RAT KIC15\_RAT STANDARD; PRT; 447 AA.

AC Q6IFV3;

DT 06-DEC-2005, integrated into UniProtKB/Swiss-Prot.

DT 05-JUL-2004, sequence version 1.

DT 07-FEB-2006, entry version 17.

DE keratin, type I cytoskeletal 15 (Cytokeratin-15) (CK-15) (Keratin-15)

DE (K15) (Type I keratin Kals1).

GN Name=Krt15; Synonym=Kals1;

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Muridae; Murinae; Rattus.

OX NCBI\_TaxId=10116;

RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=Brown Norway;

EX PubMed=15057822; DOI=10.1038/nature02426;

RA Gibbs R.A., Weinstock G.M., Metzker M.L., Muzny D.M., Sodergren E.J.,

RA Scherer S., Scott G., Steffen D., Worley K.C., Burch P.E., Okwuonu G.,

RA Hines S., Lewis L., Deramo C., Delgado O., Dugan-Rocha S., Miner G.,

RA Morgan M., Hawes A., Gill R., Holt R.A., Adams M.D., Amanatides P.G.,

RA Baden-Rillison H., Barnstead M., Chin S., Evans C.A., Ferrieria S.,

RA Poser C., Glodok A., Gu Z., Jennings D., Kraft C.L., Nguyen T.,

RA Pfankuch C.M., Sitter C., Sutton G.G., Venter J.C., Woodage T.,

RA Smith D., Lee H.M., Gustafson E., Cahill P., Kana A.,

RA Doucette-Stamm L., Weinstock K., Fechtel K., Weiss R.B., Dunn D.M.,

RA Green E.D., Blakesley R.W., Bouffard G.G., De Jong P.J., Jones S.,

RA Zhu B., Marra M., Schein J., Bosdet I., Fjell C., Jones S.,

RA Krzywinski M., Mathewson C., Siddiqui A., Wye N., McPherson J.,

RA Zhao S., Fraser C.M., Shetty J., Shateman S., Geer K., Chen Y.,

RA Abramson S., Nierman W.C., Havlak P.H., Chen R., Durbin K.J., Egan A.,

RA Ren Y., Song X.Z., Li B., Liu Y., Qin X., Cawley S., Cooney A.J.,

RA D'Souza L.M., Martin K., Wu J.Q., Gonzalez-Garay M.L., Jackson A.R.,

RA Kalafus K.J., McLeod M.P., Milosavljevic A., Virk D., Volkov A.,

RA Wheeler D.A., Zhang Z., Bailey J.A., Eichler E.E., Tuzov E.,

RA Birney E., Mongin E., Ureta-Vidal A., Woodward C., Zdobnov E.,

RA Bork P., Suyama M., Torrents D., Alexandersson M., Trask B.J.,

RA Young J.M., Huang H., Wang H., Xing H., Daniels S., Gietzen D.,

RA Schmidt J., Stevens K., Vitt U., Wingrove J., Camara F., Mar Alba M.,

RA April J.F., Guigo R., Smit A., Dubchak I., Rubin E.M., Couronne O.,

RA Poliakov A., Hubner N., Ganten D., Goesele C., Hummel O., Kreitler T.,

RA Lee Y.A., Monti J., Schulz H., Zindahl H., Himmelbauer H., Lehrach H.,

RA Jacob H.J., Bromberg S., Gullings-Handley J., Jensen-Seaman M.I.,

RA Kwitek A.E., Lazar J., Pasko D., Tonellato P.J., Twigger S.,

RA Ponting C.P., Duarte J.M., Rice S., Goodstadt L., Beeson S.A.,

RA Emes R.D., Winter E.E., Webber C., Brandt P., Nyakatura G.,

RA Adetobi M., Chiaromonte F., Elntski L., Eswara P., Hardison R.C.,

RA Hou M., Kolbe D., Makova K., Miller W., Nekutenko A., Riemer C.,

RA Schwartz S., Taylor J., Yang S., Zhang Y., Lindpaintner K.,

RA Andrews T.D., Caccamo M., Clamp M., Clarke L., Curwen V., Durbin R.,

RA Eyras E., Searle S.M., Cooper G.M., Batzoglu S., Brudno M., Sidow A.,

RA Stone E.A., Payseur B.A., Bourque G., Lopez-Otin C., Puente X.S.,

RA Chakrabarti K., Chatterji S., Dewey C., Pachter L., Bray N., Yap V.B.,

RA Caspi A., Tesler G., Pezner P.A., Haussler D., Roskin K.M.,

RA Baertsch R., Clawson H., Furey T.S., Hinrichs A.S., Karolchik D.,

RA Kent W.J., Rosenbloom K.R., Trumbower H., Weirauch M., Cooper D.N.,

RA Stenson P.D., Ma B., Brent M., Arumugam M., Shteynberg D.,  
RA Copley R.R., Taylor M.S., Riethman H., Mudunuri U., Peterson J.,  
RA Guyer M., Felsenfeld A., Old S., Mockrin S., Collins F.,  
RT "Genome sequence of the Brown Norway rat yields insights into  
RT mammalian evolution.";  
RL Nature 428:493-521(2004).

RL [2]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].

RC TISSUE=Prostate;

RG NIH - Mammalian Gene Collection (MGC) project;

RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.

RL [3]

RP IDENTIFICATION.

RC STRAIN=Brown Norway;

RX MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;

RT "Comprehensive analysis of keratin gene clusters in humans and

RT rodents.";

RL Eur. J. Cell Biol. 83:19-26(2004).

CC -!- SUBUNIT: Heterotetramer of two type I and two type II keratins.

CC -!- MISCELLANEOUS: There are two types of cytoskeletal and

CC microfibrillar keratin, I (acidic; 40-55kDa) [K9 to K20] and II

CC (neutral to basic 56-70kDa) [K1 to K8].

CC -!- SIMILARITY: Belongs to the intermediate filament family.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; AABR03073341; -; NOT\_ANNOTATED\_CDS; Genomic\_DNA.

DR EMBL; BC101868; AAT01869.1; -; mRNA.

DR EMBL; BK004045; DAA04479.1; -; mRNA.

DR Ensembl; ENSRNOG0000014099; Rattus norvegicus.

DR RGD; 1303044; Kals1.

DR InterPro; IPR001664; IF.

DR InterPro; IPR002957; Keratin\_I.

DR Pfam; PF00038; Filament; 1.

DR PRINTS; PR01248; TYPE1KERATIN.

DR PROSITE; PS00226; IF; 1.

KW Coiled coil; Intermediate filament.

FT CHAIN 1 447 Keratin, type I cytoskeletal 15.

FT /FTID=PRO\_0000063659.

FT REGION 1 93 Head.

FT REGION 94 402 Rod.

FT REGION 94 129 Coil 1A.

FT REGION 130 148 Linker 1.

FT REGION 149 240 Coil 1B.

FT REGION 241 260 Linker 12.

FT REGION 261 402 Coil 2.

FT REGION 403 447 Tail.

FT COMPIAS 13 92 Gly-rich.

SQ SEQUENCE 447 AA; 48870 MW; CD16F2D55F193671 CRC64;

Query Match 52.9%; Score 46; DB 1; Length 447;

Best Local Similarity 81.8%; Pred. No. 1e+02; Mismatches 0; Gaps 0;

Matches 9; Conservative 0; Indels 2;

OY 1 BIETVHNLEGG 11

Db 391 EISTVRLNLEGG 401

Search completed: July 5, 2006, 19:18:15

Job time : 188 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:44:12 ; Search time 147.667 Seconds  
(without alignments)  
49.540 Million cell updates/sec

Title: US-10-774-928A-3  
Perfect score: 87  
Sequence: 1 EIEYHNLEGGQEDF 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_8:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003Bs:\*  
8: Geneseq2004s:\*  
9: Geneseq2005s:\*  
10: geneseq2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	87	100.0	16	AAO26383	Aao26383 Psoriasis
2	87	100.0	16	ADV04408	Adv04408 Leishmani
3	87	100.0	622	ADS88365	Ads88365 Human pro
4	87	100.0	622	AEa15476	Aea15476 Human pol
5	87	100.0	622	AEb13340	Aeb13340 Mutant ke
6	75	86.2	600	ADFS5631	Adfs5631 Protein (
7	53	60.9	152	ABRS8600	Abrs8600 Human can
8	52.5	60.3	96	AAm21022	Aam21022 Peptide #
9	52.5	60.3	96	ABb43190	Abb43190 Peptide #
10	52.5	60.3	96	AAm37030	Aam37030 Peptide #
11	52.5	60.3	96	AAm76923	Aam76923 Human bon
12	52.5	60.3	96	ABG58585	Abg58585 Human liv
13	52.5	60.3	96	ABG46033	Abg46033 Human pep
14	52.5	60.3	148	ABO59061	Abos9061 Human gen
15	52.5	60.3	385	ABG27716	Abg27716 Novel hum
16	52.5	60.3	417	AAE33677	Aae33677 Human str
17	47	54.0	433	ABM81838	Abm81838 Tumour-as
18	47	54.0	757	ABG06703	Abg06703 Novel hum
19	47	54.0	887	ABG23684	Abg23684 Novel hum
20	47	54.0	1113	ABG06702	Abg06702 Novel hum
21	46	52.9	209	ABG16549	Abg16549 Novel hum
22	46	52.9	209	ABG16692	Abg16692 Novel hum
23	46	52.9	387	AEA20209	Aea20209 Novel hum

24	46	52.9	483	7	ADF30445	Adf30445 Mouse ang
25	46	52.9	525	9	AEC57486	Aec57486 L. acidop
26	46	52.9	529	9	AEC57318	Aec57318 L. acidop
27	45.5	52.3	32	6	ABR63670	Abr63670 Human bre
28	45.5	52.3	56	8	ADP56620	Adp56620 Human bre
29	45.5	52.3	85	6	ABP76029	Abp76029 Human GEN
30	45.5	52.3	85	6	ABP76183	Abp76183 Human GEN
31	45.5	52.3	100	9	ADZ88838	Adz88838 Breast sp
32	45.5	52.3	165	4	AAG75797	Aag75797 Human col
33	45.5	52.3	194	8	ADP56616	Adp56616 Human bre
34	45.5	52.3	256	4	AAb71659	Aab71659 Human col
35	45.5	52.3	263	8	ADP56625	Adp56625 Human bre
36	45.5	52.3	400	6	ABP55382	Abp55382 Human col
37	45.5	52.3	400	7	ADi62966	Adi62966 Human apo
38	45.5	52.3	400	7	ADi62989	Adi62989 Human apo
39	45.5	52.3	400	7	ADi62972	Adi62972 Human apo
40	45.5	52.3	400	7	ADi63032	Adi63032 Human apo
41	45.5	52.3	400	7	ADi63033	Adi63033 Human apo
42	45.5	52.3	400	7	ADi62990	Adi62990 Human apo
43	45.5	52.3	400	7	ADi63057	Adi63057 Human apo
44	45.5	52.3	400	7	ADi63010	Adi63010 Human apo
45	45.5	52.3	400	8	ADK70458	Adk70458 Respirato

ALIGNMENTS

RESULT 1	AAO26383	
ID	AAO26383	standard; peptide; 16 AA.
XX	AAO26383;	
AC	AAO26383;	
XX	XX	
DT	30-JAN-2003	(first entry)
DE	Psoriasis treating immunotherapeutic peptide, SEQ ID No 3.	
XX	XX	
KW	Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;	
KW	vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;	
KW	gene therapy.	
XX	XX	
OS	Leishmania sp.	
XX	XX	
PN	WO200274239-A2.	
XX	XX	
PD	26-SEP-2002.	
XX	XX	
PF	04-MAR-2002; 2002WO-US006496.	
XX	XX	
PR	16-MAR-2001; 2001US-00809003.	
XX	(AKIV-) AKIVA LLC.	
PI	O'daly JA;	
XX	XX	
DR	WPI; 2003-018763/01.	
XX	XX	
PT	New polypeptide and immunogenic variants comprising amino acid sequences	
PT	of particulate antigens, useful for the treatment and clinical remission	
PT	of psoriasis.	
XX	XX	
PS	Claim 1; Page 43; 56pp; English.	
XX	XX	
CC	The invention relates to a polypeptide comprising an isolated amino acid	
CC	sequence or immunogenic variants selected from any of 14 fully defined	
CC	sequences of 7-16 amino acids, given in the specification. The	
CC	immunotherapeutic agents and a pharmaceutical compositions comprising	
CC	polynucleotides and vectors of the invention are useful for the treatment	
CC	and clinical remission of psoriasis. The isolated nucleic acids are	
CC	useful as probes. The sequences of the invention can be used in the	
CC	treatment of disorders by gene therapy. This sequence represents one of	
CC	the 14 immunotherapeutic peptides of the invention	
XX	XX	

```
SQ      Sequence 16 AA;
Query Match      100.0%; Score 87; DB 6; Length 16;
Best Local Similarity 100.0%; Pred. No. 6.6e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EIETHNLLGGQEDF 16
        |||||
DB      1 EIETHNLLGGQEDF 16
        |||||

RESULT 2
ADV04408
ID      ADV04408 standard; peptide; 16 AA.
XX
AC      ADV04408;
XX
DT      24-FEB-2005 (first entry)
XX
DE      Leishmania immunotherapeutic peptide SEQ ID NO:3.
XX
KW      T-lymphocyte; immunogenicity; antigen; antiprosoriatic; immunostimulant;
KW      psoriasis.
XX
OS      Leishmania sp.
XX
PN      US2004241168-A1.
XX
PD      02-DEC-2004.
XX
PF      09-FEB-2004; 2004US-00774928.
XX
PR      16-MAR-2001; 2001US-00809003.
PR      17-OCT-2003; 2003US-00687892.
XX
PA      (ODAL/) ODALY J A.
PI      Odaly JA;
PI      WPI; 2005-011563/01.
XX
PT      Inhibiting selectively T-cell rolling in human, by administering compound
PT      interfering with cutaneous lymphocyte antigen-E selectin interaction and
PT      leukocyte function associated antigen-1 interaction.
XX
PS      Claim 9; SEQ ID NO 3; 21pp; English.
XX
CC      The invention relates to a novel method for selectively inhibiting T-cell
CC      rolling in a human host, comprising administering a compound that
CC      selectively interferes with the cutaneous lymphocyte antigen (CLA)-E
CC      selectin interaction and leukocyte function associated antigen (LFA)-
CC      1/intercellular adhesion molecule (ICAM) and very late antigen
CC      (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound
CC      includes an immunotherapeutic agent, which comprises a purified protein
CC      extract that is isolated by diethylaminoethyl Sephadex chromatography of
CC      a Nonidet P-40 insoluble particulate antigen fraction derived from
CC      isolated killed cells of anastigotes from one or more species of the
CC      Leishmania genus, where the particulate antigen fraction is solubilized
CC      with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied
CC      to diethylaminoethyl Sephadex and is eluted with a solution comprising
CC      0.1 M sodium chloride, 8 M urea and 0.025 M
CC      Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract
CC      includes polypeptides having apparent molecular weights of 73, 80 and 82
CC      kDa, after total reduction and alkylation. The species is L. amazonensis,
CC      L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the
CC      invention has antiprosoriatic and immunostimulant activity, and inhibits T-
CC      cell rolling by interfering with CLA-E selectin interaction and LFA-
CC      1/ICAM and VLA/VCAM interactions. The method is useful for selectively
CC      inhibiting T-cell rolling in a human host, and for treating psoriasis.
CC      The present sequence represents an Leishmania peptide used in the
CC      invention.
XX
SQ      Sequence 16 AA;
Query Match      100.0%; Score 87; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 6.6e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EIETHNLLGGQEDF 16
        |||||
DB      1 EIETHNLLGGQEDF 16
        |||||

RESULT 3
ADS88365
ID      ADS88365 standard; protein; 622 AA.
XX
AC      ADS88365;
XX
DT      18-NOV-2004 (first entry)
XX
DE      Human protein of a TNF-alpha signalling pathway protein complex Seq 220.
XX
KW      protein complex; tumour necrosis factor-alpha signalling pathway;
KW      TNF-alpha; chronic inflammatory disease; rheumatoid arthritis;
KW      inflammatory bowel disease; infectious disease; septic shock;
KW      bacterial infection; neurological disease; stroke-induced inflammation;
KW      neurodegenerative disease; cancer; antiinflammatory; antiarthritic;
KW      antirheumatic; cytostatic; antibacterial; gene therapy; human.
XX
OS      Homo sapiens.
XX
PN      WO2004035783-A2.
XX
PD      29-APR-2004.
XX
PF      24-SEP-2003; 2003WO-EP050655.
PR      26-SEP-2002; 2002EP-00021809.
PR      10-FEB-2003; 2003EP-00100274.
XX
PA      (CELL-) CELLZONE AG.
PI      Bouwmeester T, Huhse B, Bauch A, Ruffner H, Bauer A, Kuester B;
PI      Superti-Furga G, Kruse U;
XX
DR      WPI; 2004-348460/32.
XX
PT      New protein complex comprising at least one first and second protein of
PT      the Tumor Necrosis Factor-alpha(TNF-alpha)-signaling pathway, useful for
PT      diagnosing or treating inflammation, neurological diseases, infectious
XX
PS      Example; SEQ ID NO 220; 1980pp; English.
XX
CC      This invention relates to novel protein complexes of the tumour necrosis
CC      factor-alpha (TNF-alpha) signalling pathway. Specifically, it refers to
CC      methods for preparing these complexes comprising at least two component
CC      proteins, as well as screening methods to identify modulators of the
CC      pathway, which include antibodies, agonists and antagonists thereof. The
CC      present invention describes a protein complex and kit that are useful for
CC      diagnosing, prognosing or treating chronic inflammatory diseases such as
CC      rheumatoid arthritis and bacterial infections; infectious diseases such
CC      as stroke-induced inflammation in neurons; neurodegenerative diseases
CC      such as septic shock and inflammatory bowel disease; infectious diseases
CC      cancer. Accordingly, these complexes can be used for the development of
CC      pharmaceutical compositions that exhibit antiinflammatory, antiarthritic,
CC      antirheumatic, cytostatic and antibacterial activities and can be used
CC      for gene therapy purposes. In particular, the invention further provides
CC      siRNA-oligonucleotides useful for inhibiting protein expression for in
CC      vitro or cell culture assays. This polypeptide is a human protein that
CC      can be used in combination with other proteins provided in the
CC      specification to form novel complexes of the TNF-alpha signalling pathway
CC      of the invention.
XX
SQ      Sequence 622 AA;
```

Query Match 100.0%; Score 87; DB 8; Length 622;  
Best Local Similarity 100.0%; Pred. No. 5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQEDF 16  
DB 449 EIETHNLLGGQEDF 464

RESULT 4  
AEA15476  
ID AEA15476 standard; protein; 622 AA.

AC AEA15476;  
XX  
XX  
DT 28-JUL-2005 (first entry)  
DE Human polypeptide #95.

XX Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
KW cytostatic; neoplasm.

XX Homo sapiens.  
OS  
XX WO2005047534-A2.

PN 26-MAY-2005.  
XX  
XX 15-OCT-2004; 2004WO-EP011599.  
XX  
XX 28-OCT-2003; 2003EP-00024565.

XX (FARB ) BAYER HEALTHCARE AG.

XX Wirtz R, Munnes M;  
PI  
XX WPI; 2005-372393/38.  
DR N-PSDB; AEA15398.

XX Predicting a response to cancer treatment by detecting at least 2  
PT markers, which are genes or genomic nucleic acid sequences that are  
PT located on one chromosomal region, which is altered in malignant  
PT neoplasia.

XX Claim 7; SEQ ID NO 438; 464pp; English.

XX The invention relates to a method of predicting response to cancer  
XX treatment comprising detection of at least 2 markers, where the markers  
XX are genes and fragments or genomic nucleic acid sequences that are  
XX located on one chromosomal region, which is altered in malignant  
XX neoplasia. The invention also relates to a method for the prediction,  
XX diagnosis or prognosis of malignant neoplasia, methods for detecting  
XX deregulations in malignant neoplasia and breast cancer, a method of  
XX determining the phenotype of a cell or tissue, a method for identifying  
XX genomic regions which are altered on the chromosomal level and encode  
XX genes that are linked by function and are differentially expressed in  
XX malignant neoplasia and breast cancer, methods of screening for agents  
XX which regulate the activity of a polypeptide or a polynucleotide and  
XX antibodies that specifically bind to a full length or partial  
XX polypeptide. The method is useful for predicting response to cancer  
XX treatment. The methods and compositions are useful for predicting,  
XX diagnosing, prognosing, preventing or treating malignant neoplasia  
XX including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
XX esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
XX lung cancer. This sequence represents a human polypeptide used in the  
XX scope of the invention.

XX Sequence 622 AA;

Query Match 100.0%; Score 87; DB 9; Length 622;  
Best Local Similarity 100.0%; Pred. No. 5e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQEDF 16  
DB 449 EIETHNLLGGQEDF 464

RESULT 5  
AEB13340  
ID AEB13340 standard; protein; 622 AA.

XX AEB13340;  
AC  
XX  
XX  
DT 08-SEP-2005 (first entry)  
DE Mutant keratin-9 SEQ ID NO 5.

XX antigen; tumor-associated antigen; diagnostic; therapeutic; tumor marker;  
KW squamous cell carcinoma;  
KW autoantibody-mediated identification of antigens; AMIDA; cytostatic;  
KW neoplasm; cancer; skin tumor; ear, nose, throat disease; mutant;  
KW keratin-9.

XX Homo sapiens.  
OS  
XX Synthetic.

XX WO2005061537-A2.  
PN  
XX  
XX 07-JUL-2005.

XX 22-DEC-2004; 2004WO-EP014634.  
XX  
XX 22-DEC-2003; 2003DE-01060456.

XX (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.  
XX (VAEC-) VAECGENE BIOTECH GMBH.

XX Gires O, Zeidler R, Muenz M, Schaffrik M;  
PI  
XX WPI; 2005-479443/48.

XX Antigen for skin tumors in the nose and throat, uses at least one protein  
PT from a selected sequence or their variants with a hybridized isolated  
PT nucleic acid encoding the transcribed product.

XX Claim 1; SEQ ID NO 5; 54pp; German.

XX This invention describes the detection of novel antigens associated with  
XX squamous cell carcinomas using autoantibody-mediated identification of  
XX antigens (AMIDA). The invention also describes methods for producing  
XX antigen-presenting cells and T cells that are specific for the antigens  
XX described. The antigens can be used in a diagnostic formulation or as a  
XX pharmaceutical product for therapy or as a vaccine, for diagnosis and  
XX therapy of squamous cell tumors and skin tumors associated with the ear,  
XX nose and throat (ENT). The tumor antigen is an improved diagnostic marker  
XX for tumor identification, and is also an effective therapeutic agent.  
XX This sequence represents the tumor-associated antigen corresponding to a  
XX mutant keratin-9.

XX Sequence 622 AA;

Query Match 100.0%; Score 87; DB 9; Length 622;  
Best Local Similarity 100.0%; Pred. No. 5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQEDF 16  
DB 449 EIETHNLLGGQEDF 464

RESULT 6  
ADF55631  
ID ADF55631 standard; protein; 600 AA.

```

XX ADR55631;
AC
XX
XX 12-FEB-2004 (first entry)
XX
XX Protein (SEQ ID 1) related to oyster pearl keratin protein.
XX
XX pearl protein composite; PPC; mother of pearl manufacture; jewellery;
KW nanotechnology; oyster; keratin.
XX
XX Unidentified.
OS
XX
XX JP2003012696-A.
PN
XX
XX 15-JAN-2003.
PD
XX
XX 03-JUL-2001; 2001JP-00202033.
PF
XX
XX 03-JUL-2001; 2001JP-00202033.
PR
XX
XX (MATS/) MATSUSHIRO A.
PA
XX
XX WPI; 2003-451748/43.
DR
XX
XX Pearl protein composite useful in manufacture of mother of pearl,
PT comprises at least two protein components having preset molecular weight
PT as determined by sodium dodecyl sulfate gel electrophoresis.
XX
XX Example 4; SEQ ID NO 1; 25pp; Japanese.
XX
XX The invention relates to a novel pearl protein composite (PPC) comprising
CC at least two protein components as determined by sodium dodecyl sulphate
CC (SDS) gel electrophoresis. The composite of the invention may be useful
CC for the manufacture of mother of pearl for jewellery and also in
CC nanotechnology. The pearl protein composite (PPC) is effectively utilised
CC in manufacture of artificial pearls. The current sequence is that of the
CC protein (SEQ ID 1) of the invention which is related to oyster pearl
CC keratin protein and is homologous to a human keratin.
XX
XX
SQ Sequence 600 AA;
Query Match 86.2%; Score 75; DB 7; Length 600;
Best Local Similarity 87.5%; Pred. No. 0.00059;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQEDF 16
Db 449 EIETHNKKGGQEDF 464

RESULT 7
ABR58600
ID ABR58600 standard; protein; 152 AA.
XX
XX ABR58600;
AC
XX
XX 09-JUL-2003 (first entry)
XX
XX Human cancer related protein SEQ ID NO:257.
XX
XX Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
KW heart disease; atherosclerosis; endometriosis.
XX
XX Homo sapiens.
OS
XX
XX WO2003025138-A2.
PN
XX
XX 27-MAR-2003.
PD
XX
XX 17-SEP-2002; 2002WO-US029560.
PF
XX
XX 17-SEP-2001; 2001US-0323469P.
PR 20-SEP-2001; 2001US-0323887P.

```

```

PR 13-NOV-2001; 2001US-0350666P.
PR 08-FEB-2002; 2002US-0355145P.
PR 08-FEB-2002; 2002US-0355257P.
PR 12-APR-2002; 2002US-0372246P.
XX
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;
PI Zlounlk A;
XX
XX WPI; 2003-354600/33.
DR N-PSDB; ACC72736.
XX
XX New genes that are up-regulated or down-regulated in cancers, useful as
PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
PT therapeutic targets for screening drugs for treating these diseases.
XX
XX Claim 12; Page 745; 767pp; English.
XX
XX The present invention describes an isolated nucleic acid molecule, which
CC comprises the sequence of any of the genes that are up-regulated or down-
CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer
CC related gene nucleotide sequences which encode the proteins given in
CC ABR58521 to ABR58709. Also described: (1) determining the presence or
CC absence of a pathological cell in a patient; (2) an expression vector
CC comprising a nucleic acid molecule described above; (3) a host cell
CC comprising the vector; (4) an isolated polypeptide, which is encoded by
CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
CC of (4); (6) specifically targeting a compound to a pathological cell in a
CC patient by administering to the patient the antibody above; and (7) a
CC drug screening assay. The nucleic acid is useful as diagnostic markers or
CC therapeutic targets. In particular, the nucleic acid is useful for
CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,
CC atherosclerosis and endometriosis. The nucleic acid is also useful in
CC drug screening, particularly for identifying agents for treating these
CC pathologies
XX
XX
SQ Sequence 152 AA;
Query Match 60.9%; Score 53; DB 6; Length 152;
Best Local Similarity 75.0%; Pred. No. 0.78;
Matches 12; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 EIETHNLL--EGGQE 14
Db 129 EIETHYRLGEGGSE 144

RESULT 8
AAM21022
ID AAM21022 standard; protein; 96 AA.
XX
XX AAM21022;
AC
XX
XX 12-OCT-2001 (first entry)
XX
XX Peptide #7456 encoded by probe for measuring cervical gene expression.
DE
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
XX Homo sapiens.
OS
XX
XX WO200157278-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000670.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR

```

```
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000US-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 25848; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
XX (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human Hela cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 96 AA;
XX Query Match 60.3%; Score 52.5; DB 4; Length 96;
XX Best Local Similarity 75.0%; Pred. No. 0.56;
XX Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
XX QY 1 EITYHNLLGGQEDF 16
XX ||||| ||||| |||||
XX 42 EATYHRLLEDG-EDF 56
XX Db
XX RESULT 9
XX ABB43190
XX ID ABB43190 standard; peptide; 96 AA.
XX AC ABB43190;
XX XX
XX 04-FEB-2002 (first entry)
XX DE Peptide #10696 encoded by human foetal liver single exon probe.
XX XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX KW Homo sapiens.
XX OS
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000US-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
```

```
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 27; SEQ ID NO 35825; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 96 AA;
XX Query Match 60.3%; Score 52.5; DB 4; Length 96;
XX Best Local Similarity 75.0%; Pred. No. 0.56;
XX Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
XX QY 1 EITYHNLLGGQEDF 16
XX ||||| ||||| |||||
XX 42 EATYHRLLEDG-EDF 56
XX Db
XX RESULT 10
XX AAM37030
XX ID AAM37030 standard; protein; 96 AA.
XX AC AAM37030;
XX XX
XX 17-OCT-2001 (first entry)
XX DE Peptide #11067 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX XX
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000US-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 37299; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
```

```
SQ Sequence 96 AA;
Query Match 60.3%; Score 52.5; DB 4; Length 96;
Best Local Similarity 75.0%; Pred. No. 0.56;
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 EIETHNLLGGQEDF 16
DB 42 EIATYHRLLEDG-EDF 56

RESULT 11
AAM76923
ID AAM76923 standard; protein; 96 AA.
AC AAM76923;
XX
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 37229.
DE Human bone marrow expressed exon; gene expression analysis; probe;
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
OS
XX WO200157276-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US000668.
PF
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207458P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488900/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Claim 4; SEQ ID NO 37229; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in the human
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
XX Sequence 96 AA;
Query Match 60.3%; Score 52.5; DB 4; Length 96;
Best Local Similarity 75.0%; Pred. No. 0.56;
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 EIETHNLLGGQEDF 16
DB 42 EIATYHRLLEDG-EDF 56

RESULT 12
ABG58585
ID ABG58585 standard; peptide; 96 AA.
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW
```

```
AC ABG58585;
XX 25-FEB-2003 (first entry)
XX Human liver peptide, SEQ ID NO 37233.
XX
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX
XX Homo sapiens.
OS
XX WO200157273-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US000664.
PF
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488998/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 37233; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SEN) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG5930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 96 AA;
Query Match 60.3%; Score 52.5; DB 4; Length 96;
Best Local Similarity 75.0%; Pred. No. 0.56;
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 EIETHNLLGGQEDF 16
DB 42 EIATYHRLLEDG-EDF 56

RESULT 13
ABG46033
ID ABG46033 standard; peptide; 96 AA.
XX
XX AC ABG46033;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 35698.
XX
XX KW
```





CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterizing  
CC alternative splicing events, in detecting and characterizing gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe protein of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX

SQ Sequence 148 AA;

Query Match 60.3%; Score 52.5; DB 8; Length 148;  
Best Local Similarity 75.0%; Pred. No. 0.93;  
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 EIETVHNLEGGQEDF 16

Db 101 EITTYHRLLEDG-EDF 115

RESULT 15

ABG27716

ID ABG27716 standard; protein; 385 AA.

XX

AC ABG27716;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #27707.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

FN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

XX

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

XX WPI; 2001-639362/73.

DR

DR N-PSDB; AAS91903.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX

PS Claim 20; SEQ ID NO 58075; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 385 AA;

Query Match 60.3%; Score 52.5; DB 4; Length 385;

Best Local Similarity 75.0%; Pred. No. 2.9;  
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 EIETVHNLEGGQEDF 16

Db 337 EITTYHRLLEDG-EDF 351

Search completed: July 5, 2006, 19:08:51

Job time : 149.667 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:52:16 ; Search time 84.6667 Seconds  
(without alignments)  
87.537 Million cell updates/sec

Title: US-10-774-928A-3

Perfect score: 87

Sequence: 1 E1ETHNLLGGQEDF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	16	5	US-10-774-928-3
2	52.5	60.3	96	3	Sequence 3, Appli
3	52.5	60.3	148	4	Sequence 47444, A
4	52.5	60.3	385	5	Sequence 32695, A
5	52.5	60.3	417	5	Sequence 58075, A
6	47	54.0	294	4	Sequence 11, Appl
7	47	54.0	887	5	Sequence 253404, A
8	47	54.0	1113	5	Sequence 37062, A
9	46	52.9	209	5	Sequence 54043, A
10	46	52.9	209	5	Sequence 37061, A
11	46	52.9	209	5	Sequence 46908, A
12	46	52.9	483	4	Sequence 47051, A
13	45.5	52.3	85	5	Sequence 8, Appli
14	45.5	52.3	85	5	Sequence 236, App
15	45.5	52.3	165	4	Sequence 509, App
16	45.5	52.3	286	3	Sequence 6571, Ap
17	45.5	52.3	400	3	Sequence 36, Appl
18	45.5	52.3	400	3	Sequence 1115, Ap
19	45.5	52.3	400	4	Sequence 1115, Ap
20	45.5	52.3	400	5	Sequence 118, Appl
21	45.5	52.3	400	5	Sequence 33, Appl
22	45.5	52.3	400	6	Sequence 1115, Ap
23	45.5	52.3	428	5	Sequence 10382, A
24	45	51.7	1184	4	Sequence 40382, A
25	44.5	51.1	7465	6	Sequence 183853, A
26	44.5	51.1	130	5	Sequence 7521, Ap
27	44.5	51.1	130	5	Sequence 44463, A
					Sequence 44564, A
					Sequence 44828, A

28	44.5	51.1	130	5	US-10-450-763-48987	Sequence 48987, A
29	44.5	51.1	232	3	US-09-919-497-78	Sequence 78, Appl
30	44.5	51.1	332	5	US-10-450-763-49052	Sequence 49052, A
31	44.5	51.1	359	5	US-10-883-020-6	Sequence 6, Appli
32	44.5	51.1	364	5	US-10-450-763-58076	Sequence 58076, A
33	44.5	51.1	375	4	US-10-080-334-268	Sequence 268, App
34	44.5	51.1	423	4	US-10-080-334-269	Sequence 269, App
35	44.5	51.1	423	5	US-10-514-932-2	Sequence 2, Appli
36	44.5	51.1	424	4	US-10-080-334-266	Sequence 266, App
37	44.5	51.1	429	5	US-10-883-020-2	Sequence 2, Appli
38	44.5	51.1	430	3	US-09-919-039-105	Sequence 105, App
39	44.5	51.1	430	3	US-09-813-432-36	Sequence 36, Appl
40	44.5	51.1	430	4	US-10-026-001-1	Sequence 1, Appli
41	44.5	51.1	430	4	US-10-060-036-155	Sequence 155, App
42	44.5	51.1	430	4	US-10-174-364-36	Sequence 36, Appl
43	44.5	51.1	430	4	US-10-080-334-265	Sequence 265, App
44	44.5	51.1	430	4	US-10-080-334-267	Sequence 267, App
45	44.5	51.1	430	4	US-10-246-583-36	Sequence 36, Appl

## ALIGNMENTS

RESULT 1  
US-10-774-928-3  
; Sequence 3, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 3  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-3

Query Match 100.0%; Score 87; DB 5; Length 16;  
Best Local Similarity 100.0%; Pred. No. 4.8e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 E1ETHNLLGGQEDF 16  
|||  
DB 1 E1ETHNLLGGQEDF 16

RESULT 2  
US-09-864-761-47444  
; Sequence 47444, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04

```

/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/006666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annonax Sequence Listing Engine
/ SEQ ID NO 47444
/ LENGTH: 96
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC008739.5
/ OTHER INFORMATION: EXPRESSED IN PLACENT
/ OTHER INFORMATION: EXPRESSED IN FETAL L
/ OTHER INFORMATION: EXPRESSED IN BONE MAR
/ OTHER INFORMATION: EXPRESSED IN LUNG, S
/ OTHER INFORMATION: EXPRESSED IN ADULT L
/ OTHER INFORMATION: EXPRESSED IN HELA, S
/ OTHER INFORMATION: SWISSPROT HIT: P0578
/ OTHER INFORMATION: EST_HUMAN HIT: AA829
/ US-09-864-761-47444

```

```
Query Match          60.3%; Score 52.5; DB 3; Length 96;
Best Local Similarity 75.0%; Pred. No. 1.3;
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
```

Qy 1 EIETHNLLGGQEDF 16  
||| ||| ||| ||| |||  
Db 42 EIATYHRLLEDG-EDF 56

```

RESULT 3
US-10-029-386-32695
; Sequence 32695, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED
; TITLE OF INVENTION: EXPRESSION ANALYSIS
; FILE REFERENCE: ACOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine
; SEQ ID NO 32695
; LENGTH: 148
; TYPE: PRT

```

```

? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC000094.3
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.72
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7
? OTHER INFORMATION: SWISSPROT HIT: P05783, EVALUAE 1.00e-57
? US-10-029-386-32695

```

```

Query Match      60.3%; Score 52.5; DB 4; Length 148;
Best Local Similarity 75.0%; Pred. No. 2;
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

Qy	1	EI	TY	HN	LE	GG	QD	F	16
Db	101	EI	TY	HR	LE	DG	-E	D	119

```

RESULT 4
US-10-450-763-58075
; Sequence 58075, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hvaeq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 58075
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (306)..(353)
; OTHER INFORMATION: Intermediate filaments proteins domain
; OTHER INFORMATION: accession number BL00226D, p-value=8
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (169)..(352)
; OTHER INFORMATION: Intermediate filament proteins domain
; OTHER INFORMATION: accession name filament, E-value=5.59
US-10-450-763-58075

```

```
Query Match          60.3%; Score 52.5; DB 5; Length 385;
Best Local Similarity 75.0%; Pred. No. 5.5;
Matches 12; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
```

Qy 1 EIETHNLEGGQEDF 16  
||| ||| ||| |||  
Db 337 EITTYHRLLEDG-EDF 351

RESULT 5  
US-10-479-764-11  
; Sequence 11, Application US/10479764  
; Publication No. US20050027103A1  
; GENERAL INFORMATION:  
; APPLICANT: TANG, Y. Tom; WARREN, Bridget A.;  
; APPLICANT: HONCHELL, Cynthia D.; RICHARDSON, Thomas W.;  
; APPLICANT: ELLIOTT, Vicki S.; CHAWLA, Narinder K.;  
; APPLICANT: YUE, Henry; BATRA, Sajeev;  
; APPLICANT: GRIFFIN, Jennifer A.; BAUGHN, Mariah R.;



; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 54043  
; LENGTH: 887  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (547)..(567)  
; OTHER INFORMATION: kw ALLERGEN POLLEN CIM1 HOL-II domain identified by eMATRIX,  
; OTHER INFORMATION: accession number DM01724, p-value=8.091e-11, raw score of 8.14  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (84)..(857)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
; OTHER INFORMATION: accession name filament, E-value=2e-247, Pfam score of 835.4  
US-10-450-763-54043

Query Match 54.0%; Score 47; DB 5; Length 887;  
Best Local Similarity 81.8%; Pred. No. 1e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0

Qy 1 EIETHNLLLEG 11  
||| ||| ||| |||  
Db 394 EITVHRLLEG 404

## RESULT 9

US-10-450-763-37061  
; Sequence 37061, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 37061  
; LENGTH: 1113  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1004)..(1024)  
; OTHER INFORMATION: kw ALLERGEN POLLEN CIM1 HOL-II domain identified by eMATRIX,  
; OTHER INFORMATION: accession number DM01724, p-value=3.045e-11, raw score of 8.14  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (84)..(1104)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
; OTHER INFORMATION: accession name filament, E-value=3.3e-277, Pfam score of 934.3  
US-10-450-763-37061

Query Match 54.0%; Score 47; DB 5; Length 1113;  
Best Local Similarity 81.8%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0

Qy 1 EIETHNLLLEG 11  
||| ||| ||| |||  
Db 394 EITVHRLLEG 404

## RESULT 10

US-10-450-763-46908

; Sequence 46908, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 46908  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (124)..(171)  
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,  
; OTHER INFORMATION: accession number BL00226D, p-value=3.250e-25, raw score of 19.10  
US-10-450-763-46908

Query Match 52.9%; Score 46; DB 5; Length 209;  
Best Local Similarity 64.3%; Pred. No. 33; Mismatches 2; Indels 0; Gaps 0;  
Matches 9; Conservative 2

Qy 1 EIETHNLLLEGQ 14  
||| ||| ||| |||  
Db 155 EITVHCLLEGDKD 168

## RESULT 11

US-10-450-763-47051  
; Sequence 47051, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 47051  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (124)..(171)  
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,  
; OTHER INFORMATION: accession number BL00226D, p-value=3.250e-25, raw score of 19.10  
US-10-450-763-47051

Qy 1 EIETHNLLLEG 11  
||| ||| ||| |||  
Db 394 EITVHRLLEG 404

US-10-450-763-47051

Query Match 52.9%; Score 46; DB 5; Length 209;  
Best Local Similarity 64.3%; Pred. No. 33;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQE 14  
||| ||| ||| ||| :  
Db 155 EITTHCLLEDGKD 168

RESULT 12

US-10-316-253-8  
; Sequence 8, Application US/10316253  
; Publication No. US20030162706A1  
; GENERAL INFORMATION:  
; APPLICANT: The Procter & Gamble Company  
; APPLICANT: Peters, Kevin  
; APPLICANT: Thompson, Larry  
; APPLICANT: Wang, Feng  
; APPLICANT: Greis, Kenneth  
; TITLE OF INVENTION: Angiogenesis Modulating Proteins  
; FILE REFERENCE: 8865M  
; CURRENT APPLICATION NUMBER: US/10/316,253  
; PRIOR FILING DATE: 2002-12-10  
; PRIOR APPLICATION NUMBER: US 60/355,295  
; PRIOR FILING DATE: 2002-02-08  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 8  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-316-253-8

Query Match 52.9%; Score 46; DB 4; Length 483;  
Best Local Similarity 64.3%; Pred. No. 80;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 EIETHNLLGGQE 14  
||| ||| ||| ||| :  
Db 418 EITTHRLLEDGQ 431

RESULT 13

US-10-475-075-236  
; Sequence 236, Application US/10475075  
; Publication No. US20060053498A1  
; GENERAL INFORMATION:  
; APPLICANT: Bejanin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Jobert, Severin  
; APPLICANT: Giordano, Jean-Yves  
; TITLE OF INVENTION: Full-length human cDNAs encoding potentially secreted proteins  
; FILE REFERENCE: G-081US03PCT  
; CURRENT APPLICATION NUMBER: US/10/475,075  
; CURRENT FILING DATE: 2003-10-17  
; PRIOR APPLICATION NUMBER: PCT/IB01/00914  
; PRIOR FILING DATE: 2001-04-18  
; NUMBER OF SEQ ID NOS: 918  
; SOFTWARE: Patent.pm  
; SEQ ID NO 236  
; LENGTH: 85  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -32..-1  
US-10-475-075-236

Query Match 52.3%; Score 45.5; DB 5; Length 85;  
Best Local Similarity 73.3%; Pred. No. 15;

Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQE 15  
||| ||| ||| ||| :  
Db 61 EIATYRSLLJL-GQED 74

RESULT 14

US-10-475-075-509  
; Sequence 509, Application US/10475075  
; Publication No. US20060053498A1  
; GENERAL INFORMATION:  
; APPLICANT: Bejanin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Jobert, Severin  
; APPLICANT: Giordano, Jean-Yves  
; TITLE OF INVENTION: Full-length human cDNAs encoding potentially secreted proteins  
; FILE REFERENCE: G-081US03PCT  
; CURRENT APPLICATION NUMBER: US/10/475,075  
; CURRENT FILING DATE: 2003-10-17  
; PRIOR APPLICATION NUMBER: PCT/IB01/00914  
; PRIOR FILING DATE: 2001-04-18  
; NUMBER OF SEQ ID NOS: 918  
; SOFTWARE: Patent.pm  
; SEQ ID NO 509  
; LENGTH: 85  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -32..-1  
US-10-475-075-509

Query Match 52.3%; Score 45.5; DB 5; Length 85;

Best Local Similarity 73.3%; Pred. No. 15;  
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQE 15  
||| ||| ||| ||| :  
Db 61 EIATYRSLLJL-GQED 74

RESULT 15

US-10-106-698-6571  
; Sequence 6571, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: Patent in Ver. 3.0  
; SEQ ID NO 6571  
; LENGTH: 165  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (4)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (6)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE

; LOCATION: (9)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC FEATURE  
; LOCATION: (22)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC FEATURE  
; LOCATION: (25)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-106-698-6571

Query Match 52.3%; Score 45.5; DB 4; Length 165;  
Best Local Similarity 73.3%; Pred. No. 31;  
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 EIETHNLLGGQED 15  
|||:|||||  
Db 141 EIATYRSLLE-QQED 154

Search completed: July 5, 2006, 20:57:19  
Job time : 85.6667 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:53:13 ; Search time 7.3333 Seconds  
(without alignments)  
58.543 Million cell updates/sec

Title: US-10-774-928A-3

Perfect score: 87

Sequence: 1 EIETHNLLGGQEDF 16

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45.5	52.3	400	7	US-11-105-233-182
2	44.5	51.1	430	7	US-11-105-233-181
3	42	48.3	209	7	US-11-274-634-19
4	42	48.3	422	7	US-11-105-233-176
5	40	46.0	172	7	US-11-293-697-4778
6	40	46.0	356	6	US-10-953-349-38914
7	40	46.0	356	7	US-11-174-307B-4934
8	40	46.0	358	6	US-10-953-349-38913
9	40	46.0	491	7	US-11-293-697-4603
10	40	46.0	586	6	US-10-505-928-314
11	40	46.0	977	6	US-10-449-902-50081
12	39	44.8	356	6	US-10-953-349-24114
13	39	44.8	403	7	US-11-270-040-2
14	39	44.8	423	6	US-10-953-349-24113
15	39	44.8	424	7	US-11-105-233-183
16	39	44.8	443	6	US-10-953-349-24112
17	39	44.8	450	7	US-11-293-697-4090
18	39	44.8	466	7	US-11-289-102-289
19	39	44.8	469	6	US-10-505-928-457
20	39	44.8	483	7	US-11-105-233-180
21	39	44.8	520	7	US-11-293-697-3712
22	39	44.8	551	7	US-11-270-040-4
23	39	44.8	864	6	US-10-449-902-41176
24	38	43.7	91	6	US-10-953-349-29533
25	38	43.7	151	6	US-10-953-349-29532

```

26 38 43.7 254 6 US-10-449-902-33071 Sequence 33071, A
27 38 43.7 309 7 US-11-178-538-55 Sequence 55, Appl
28 38 43.7 309 7 US-11-178-538-62 Sequence 62, Appl
29 38 43.7 319 6 US-10-953-349-16459 Sequence 16459, A
30 38 43.7 386 6 US-10-953-349-16458 Sequence 16458, A
31 38 43.7 390 6 US-10-953-349-16457 Sequence 16457, A
32 38 43.7 557 7 US-11-233-089-46 Sequence 46, Appl
33 38 43.7 578 6 US-10-449-902-43188 Sequence 43188, A
34 38 43.7 599 6 US-10-449-902-49534 Sequence 49534, A
35 38 43.7 929 7 US-11-165-586-55 Sequence 55, Appl
36 38 43.7 2937 6 US-10-480-962-18 Sequence 18, Appl
37 38 43.7 2969 6 US-10-480-962-19 Sequence 19, Appl
38 37 42.5 195 6 US-10-449-902-43488 Sequence 43488, A
39 37 42.5 339 6 US-10-449-902-29389 Sequence 29389, A
40 37 42.5 339 6 US-10-449-902-32420 Sequence 32420, A
41 37 42.5 401 6 US-10-449-902-56324 Sequence 56324, A
42 37 42.5 504 6 US-10-449-902-33553 Sequence 33553, A
43 37 42.5 827 6 US-10-449-902-56545 Sequence 56545, A
44 37 42.5 845 6 US-10-449-902-54753 Sequence 54753, A
45 37 42.5

```

## ALIGNMENTS

```

RESULT 1
US-11-105-233-182
; Sequence 182, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-182

```

```

Query Match 52.3%; Score 45.5; DB 7; Length 400;
Best Local Similarity 73.3%; Pred. No. 2.9;
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 1 EIETHNLLGGQED 15
DB 376 EIATYRSLLG-QGED 389

```

```

RESULT 2
US-11-105-233-181
; Sequence 181, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-181

```

```

Query Match 51.1%; Score 44.5; DB 7; Length 430;

```





```
; PRIOR APPLICATION NUMBER: 60/680,297
; PRIOR FILING DATE: 2005-05-12
; PRIOR APPLICATION NUMBER: 60/626,788
; PRIOR FILING DATE: 2004-11-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Sus scrofa
US-11-270-040-2

Query Match          44.8%; Score 39; DB 7; Length 403;
Best Local Similarity 72.7%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 EIETHNLLG 11
Db      266 EIATYRKLLG 276

RESULT 14
US-10-953-349-24113
; Sequence 24113, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 24113
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-24113

Query Match          44.8%; Score 39; DB 6; Length 423;
Best Local Similarity 70.0%; Pred. No. 39;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      4 TYHNLGGQ 13
Db      401 TINNLGGQ 410

RESULT 15
US-11-105-233-183
; Sequence 183, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thigalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 183
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-183

Query Match          44.8%; Score 39; DB 7; Length 424;
Best Local Similarity 72.7%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 EIETHNLLG 11
```

```
; PRIOR APPLICATION NUMBER: 104499502
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-Oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50081
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-50081

Query Match          46.0%; Score 40; DB 6; Length 977;
Best Local Similarity 70.0%; Pred. No. 67;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 IETHNLLG 11
Db      255 LELYNNLLG 264

RESULT 12
US-10-953-349-24114
; Sequence 24114, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 24114
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-24114

Query Match          44.8%; Score 39; DB 6; Length 356;
Best Local Similarity 70.0%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      4 TYHNLGGQ 13
Db      334 TINNLGGQ 343

RESULT 13
US-11-270-040-2
; Sequence 2, Application US/11270040
; Publication No. US20060110762A1
; GENERAL INFORMATION:
; APPLICANT: Kapil, Sanjay
; APPLICANT: Kim, Jeong-Ki
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY
; TITLE OF INVENTION: SYNDROME VIRUS RECEPTOR COMPONENTS AND USES THEREOF
; FILE REFERENCE: 14337/004001
; CURRENT APPLICATION NUMBER: US/11/270,040
; CURRENT FILING DATE: 2005-11-09
```

Db           || || || || ||  
              366 E1ATYRELLEG 376

Search completed: July 5, 2006, 20:57:47  
Job time : 7.33333 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:24:38 ; Search time 28.3333 Seconds  
(without alignments)  
49.429 Million cell updates/sec

Title: US-10-774-928A-3

Perfect score: 87  
Sequence: 1 EIETHNLLGGQEDF 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/iaa/5 COMB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/iaa/6 COMB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/iaa/7 COMB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/iaa/H COMB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/iaa/PCTUS COMB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/iaa/RE COMB.pep.\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	16	2 US-09-809-003A-3	Sequence 3, Appli
2	45.5	52.3	400	1 US-07-730-953-2	Sequence 2, Appli
3	45.5	52.3	442	2 US-09-949-016-8625	Sequence 8625, Ap
4	44.5	51.1	232	2 US-09-919-497-78	Sequence 78, Appl
5	44.5	51.1	430	2 US-09-919-039-105	Sequence 105, App
6	44.5	51.1	447	2 US-09-949-016-11032	Sequence 11032, A
7	44	50.6	728	1 US-08-475-844-5	Sequence 5, Appli
8	44	50.6	728	5 PCT-US95-08429-5	Sequence 5, Appli
9	43	49.4	323	2 US-09-248-796A-24078	Sequence 24078, A
10	43	49.4	593	2 US-09-538-092-919	Sequence 919, App
11	43	49.4	911	1 US-08-928-692-59	Sequence 59, Appl
12	43	49.4	911	2 US-09-339-972-59	Sequence 59, Appl
13	43	49.4	916	1 US-08-928-692-58	Sequence 58, Appl
14	43	49.4	916	2 US-09-339-972-58	Sequence 58, Appl
15	42	48.3	422	1 US-09-067-351-3	Sequence 3, Appli
16	42	48.3	422	2 US-09-360-490-3	Sequence 3, Appli
17	42	48.3	428	2 US-09-248-796A-14834	Sequence 14834, A
18	41	47.1	235	2 US-09-620-405B-487	Sequence 487, App
19	41	47.1	235	2 US-09-604-287A-487	Sequence 487, App
20	41	47.1	235	2 US-09-834-759-487	Sequence 487, App
21	41	47.1	235	2 US-10-076-622-487	Sequence 487, App
22	41	47.1	235	2 US-10-124-805-487	Sequence 487, App
23	41	47.1	274	2 US-09-949-016-9095	Sequence 9095, Ap
24	41	47.1	335	2 US-09-810-506-2	Sequence 2, Appli
25	41	47.1	493	2 US-09-949-016-6203	Sequence 6203, Ap
26	41	47.1	505	2 US-09-620-405B-478	Sequence 478, App

27	41	47.1	505	2 US-09-620-405B-485	Sequence 485, App
28	41	47.1	505	2 US-09-604-287A-478	Sequence 478, App
29	41	47.1	505	2 US-09-604-287A-485	Sequence 485, App
30	41	47.1	505	2 US-09-834-759-478	Sequence 478, App
31	41	47.1	505	2 US-09-834-759-485	Sequence 485, App
32	41	47.1	505	2 US-09-590-751A-478	Sequence 478, App
33	41	47.1	505	2 US-09-551-621-478	Sequence 478, App
34	41	47.1	505	2 US-09-551-621A-478	Sequence 478, App
35	41	47.1	505	2 US-10-076-622-478	Sequence 478, App
36	41	47.1	505	2 US-10-076-622-485	Sequence 485, App
37	41	47.1	505	2 US-10-124-805-478	Sequence 478, App
38	41	47.1	505	2 US-10-124-805-485	Sequence 485, App
39	41	47.1	615	1 US-08-484-105-6	Sequence 6, Appli
40	41	47.1	615	1 US-08-484-105-6	Sequence 6, Appli
41	40.5	46.6	634	2 US-10-164-595-69	Sequence 69, Appl
42	40.5	46.6	725	2 US-10-164-595-30	Sequence 30, Appl
43	40.5	46.6	1031	2 US-09-914-259-24	Sequence 24, Appl
44	40	46.0	185	2 US-09-543-681A-6449	Sequence 6449, Ap
45	40	46.0	251	2 US-09-252-991A-32188	Sequence 32188, A

## ALIGNMENTS

## RESULT 1

US-09-809-003A-3  
; Sequence 3, Application US/09809003A  
; Patent No. 6673351

; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: Psoriasis

; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16

; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.0

; SEQ ID NO 3  
; LENGTH: 16  
; TYPE: PRT

; ORGANISM: Leishmania  
US-09-809-003A-3

Query Match 100.0%; Score 87; DB 2; Length 16;  
Best Local Similarity 100.0%; Pred. No. 7.7e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	EIETHNLLGGQEDF	16
DB	1	EIETHNLLGGQEDF	16

## RESULT 2

US-07-730-953-2  
; Sequence 2, Application US/07730953  
; Patent No. 5286614

; GENERAL INFORMATION:  
; APPLICANT: BODENMULLER, Heinz

; APPLICANT: DESSAUER, Andreas  
; TITLE OF INVENTION: METHOD FOR THE DETECTION OF MALIGNANT

; TITLE OF INVENTION: DISEASES  
; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram

; STREET: 655 Fifteenth Street N.W. Suite 330  
; CITY: Washington

; STATE: D.C.  
; COUNTRY: U.S.A.

; ZIP: 20005-5701  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/07/730,953  
; FILING DATE: 19910723  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 39 42 999.7  
; FILING DATE: 21-DEC-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitts, Monica C.  
; REGISTRATION NUMBER: 36,105  
; REFERENCE/DOCKET NUMBER: P564-1119  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)638-5000  
; TELEFAX: (202)638-4810  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 400 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-730-953-2

Query Match 52.3%; Score 45.5; DB 1; Length 400;  
Best Local Similarity 73.3%; Pred. No. 19;  
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQED 15  
||| :||| ||||  
Db 376 EIATYRSLLG-QED 389

RESULT 3  
US-09-949-016-8625  
; Sequence 8625, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8625  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-8625

Query Match 52.3%; Score 45.5; DB 2; Length 442;  
Best Local Similarity 73.3%; Pred. No. 21;  
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQED 15  
||| :||| ||||  
Db 418 EIATYRSLLG-QED 431

RESULT 4  
US-09-919-497-78  
; Sequence 78, Application US/09919497  
; Patent No. 6773883  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER

; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 78  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-497-78

Query Match 51.1%; Score 44.5; DB 2; Length 232;  
Best Local Similarity 68.8%; Pred. No. 15;  
Matches 11; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQEDF 16  
||| :||| ||||  
Db 178 EIATYRRLLEDG-EDF 192

RESULT 5  
US-09-919-039-105  
; Sequence 105, Application US/09919039  
; Patent No. 6727066  
; GENERAL INFORMATION:  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
; FILE REFERENCE: PA-0035 US  
; CURRENT APPLICATION NUMBER: US/09/919,039  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: 60/222,113  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 401  
; SOFTWARE: PERL Program  
; SEQ ID NO 105  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6727066 1272969CDI  
US-09-919-039-105

Query Match 51.1%; Score 44.5; DB 2; Length 430;  
Best Local Similarity 68.8%; Pred. No. 30;  
Matches 11; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 EIETHNLLGGQEDF 16  
||| :||| ||||  
Db 376 EIATYRRLLEDG-EDF 390

RESULT 6  
US-09-949-016-11032  
; Sequence 11032, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0



```

; SEQ ID NO 11032
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11032

```

Query Match 51.1%; Score 44.5; DB 2; Length 447;  
Best Local Similarity 68.8%; Pred. No. 31;  
Matches 11; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```

RESULT 7
US-08-475-844-5
; Sequence 5, Application US/08475844
; Patent No. 5972643
; GENERAL INFORMATION:
; APPLICANT: Lobanenkov, Victor V.
; APPLICANT: Neiman, Paul E.
; APPLICANT: Klenova, Elena M.
; APPLICANT: Goodman, Graham H.
; APPLICANT: Filippova, Galina N.
; APPLICANT: Collins, Steven J.
; TITLE OF INVENTION: CTCF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA

```

```
Query Match      50.6%;      Score 44;      DB 1;      Length 728;
Best Local Similarity 60.0%;      Pred. No. 65;
Matches 9;      Conservative 1;      Mismatches 5;      Indels
```

## RESULT 8

```

PCT-US95-08429-5
; Sequence 5, Application PC/TUS9508429
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CTCF
; NUMBER OF SEQUENCES: 21
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08429
; FILING DATE: 15-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,680
; FILING DATE: 17-JUN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14538A-11-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 728 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-08429-5

```

Query Match	50.6%	Score 44;	DB 5;	Length 728;
Best Local Similarity	60.0%	Pred. No. 65;		
Matches	9;	Conservative	1;	Mismatches
			5;	Indels

```

RESULT 9
US-09-248-796A-24078
; Sequence 24078, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUEN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 24078
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-24078

```

Query Match	49.4%	Score 43;	DB 2;	Length 323;
Best-Local Similarity	53.8%	Pred. NO. 39;		
Matches	7;	Conservative	3;	Mismatches
			3;	Indels

```
RESULT 10
US-09-538-092-919
; Sequence 919, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormatter Version 0.9
; SEQ ID NO 919
; LENGTH: 593
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P13645
US-09-538-092-919

Query Match      49.4%; Score 43; DB 2; Length 593;
Best Local Similarity 72.7%; Pred. No. 76;
Matches      8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 RIETVHNLLG 11
Db      445 EIQTYSLLG 455

RESULT 11
US-08-928-692-59
; Sequence 59, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,692
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 59:
; LENGTH: 911 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6323002e
US-09-339-972-59

Query Match      49.4%; Score 43; DB 2; Length 911;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches      8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 YHNLEGGQEDF 16
Db      171 YRNDLEGGSEEF 182

RESULT 12
US-09-339-972-59
; Sequence 59, Application US/09339972
; Patent No. 6323002
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6323002o No. 6323002disk of No. 6323002th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/339,972
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,692
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 59:
; LENGTH: 911 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6323002e
US-09-339-972-59

Query Match      49.4%; Score 43; DB 2; Length 911;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches      8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 YHNLEGGQEDF 16
Db      171 YRNDLEGGSEEF 182
```

RESULT 13  
US-08-928-692-58  
; Sequence 58, Application US/08928692  
; Patent No. 5958727  
; GENERAL INFORMATION:  
; APPLICANT: Brody, Howard  
; APPLICANT: Yaver, Deborah S.  
; APPLICANT: Lamsa, Michael  
; APPLICANT: Hansen, Kim  
; TITLE OF INVENTION: Methods for Modifying the Production of  
; TITLE OF INVENTION: a Polypeptide  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5958727a No. 5958727a disk of No. 5958727th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,692  
; FILING DATE: 12-SEPT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 4944.200-US  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5958727e  
US-08-928-692-58  
Query Match 49.4%; Score 43; DB 1; Length 916;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 5 YNNLLEGGQEDF 16  
Db 176 YRNDLEGGSEEF 187  
RESULT 14  
US-09-339-972-58  
; Sequence 58, Application US/09339972  
; Patent No. 6323002  
; GENERAL INFORMATION:  
; APPLICANT: Brody, Howard  
; APPLICANT: Yaver, Deborah S.  
; APPLICANT: Lamsa, Michael  
; APPLICANT: Hansen, Kim  
; TITLE OF INVENTION: Methods for Modifying the Production of  
; TITLE OF INVENTION: a Polypeptide  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6323002a No. 6323002a disk of No. 6323002th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,351  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 422 amino acids

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/339,972  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/928,692  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 4944.200-US  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 916 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 6323002e  
US-09-339-972-58  
Query Match 49.4%; Score 43; DB 2; Length 916;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 5 YNNLLEGGQEDF 16  
Db 176 YRNDLEGGSEEF 187  
RESULT 15  
US-09-067-351-3  
; Sequence 3, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,351  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 422 amino acids

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT07
; CLONE: 2124178
US-09-067-351-3
Query Match      48.3%; Score 42; DB 1; Length 422;
Best Local Similarity 64.3%; Pred. No. 76;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 EIETYNLLEGGQE 14
          |||||
Db      367 EITTYRLLEGESE 380

Search completed: July 5, 2006, 19:26:43
Job time : 29.3333 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 10.9223 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928A-4

Perfect score: 45

Sequence: 1 AQVEDAIQK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80:\*

1: Pirl:\*

2: Pirl2:\*

3: Pirl3:\*

4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	82.2	785	2 AB1582	MutS protein (MutS)
2	35	77.8	174	2 S39790	male-enhanced anti
3	35	77.8	174	2 B34421	male-enhanced anti
4	35	77.8	185	2 A34421	male-enhanced anti
5	35	77.8	401	2 A96941	hypothetical prote
6	35	77.8	449	2 G87789	protein C34G6.7 [i
7	35	77.8	785	2 AH1228	MutS protein (MutS
8	35	77.8	875	2 S70115	Zif1 protein - yea
9	35	77.8	2327	2 T31733	hypothetical prote
10	34	75.6	416	2 T37196	hypothetical prote
11	33	73.3	142	2 T28386	hypothetical prote
12	33	73.3	356	2 T27172	hypothetical prote
13	33	73.3	477	2 A49937	50k outer membrane
14	33	73.3	485	2 F83593	outer membrane pro
15	33	73.3	591	2 T03602	probable asparagin
16	33	73.3	708	2 T18779	transposase contai
17	33	73.3	854	2 S02003	neurofilament trip
18	33	73.3	908	2 C70168	DNA polymerase I (
19	33	73.3	1072	1 A37221	neurofilament trip
20	33	73.3	1087	1 QFM5H	neurofilament trip
21	32	71.1	98	2 D64005	hypothetical prote
22	32	71.1	105	2 T10152	hypothetical prote
23	32	71.1	133	2 JC5982	preylated protein
24	32	71.1	132	2 C75220	hypothetical prote
25	32	71.1	195	2 F71202	hypothetical prote
26	32	71.1	227	2 D86631	transposase of ISI
27	32	71.1	227	2 G86704	transposase of ISI
28	32	71.1	246	2 E97161	conserved membrane
29	32	71.1	272	2 C86642	transposase of ISI

#### ALIGNMENTS

##### RESULT 1

AB1582

MutS protein (MutS2) homolog lin1195 [imported] - Listeria innocua (strain Clip11262)

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 31-Dec-2004

C:Accession: AB1582

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat

ok, C.; Schluter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AB1582

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-785 <GLA>

A:Cross-references: UNIPROT:Q92CH6; UNIPARC:UPI000012FA23; GB:AL592022; PIDN:CAC96426.1;

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin1195

C:Superfamily: DNA mismatch repair protein, bacterial/plant types

Query Match 82.2%; Score 37; DB 2; Length 785;

Best Local Similarity 87.5%; Pred. No. 18;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9

DB 716 RYEDAIQK 723

##### RESULT 2

S39790

male-enhanced antigen - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 19-May-1994 #sequence\_revision 19-Apr-1996 #text\_change 09-Jul-2004

C:Accession: S39790

R:Kondo, M.; Sato, S.; Sutou, S.

Biochim. Biophys. Acta 1216, 483-486, 1993

A:Title: Cloning and sequence analysis of cDNA encoding the bovine testis-derived male-e-

A:Reference number: S39790; MUID:94092744; PMID:8268232

A:Accession: S39790

A:Molecule type: mRNA

A:Residues: 1-174 <KON>

A:Cross-references: UNIPROT:Q29407; UNIPARC:UPI000012EE9F; EMBL:DI7340; NID:g560052; PIDN:

A>Note: it is uncertain whether Met-1 or Met-11 is the initiator

Query Match 77.8%; Score 35; DB 2; Length 174;

Best Local Similarity 66.7%; Pred. No. 8.7;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|:|:|:|:|  
Db 154 AQWEDVQK 162

RESULT 3  
B34421  
male-enhanced antigen - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Jun-1990 #sequence\_revision 09-Oct-1992 #text\_change 09-Jul-2004  
C:Accession: B34421  
R:Laue, Y.F.C.; Chan, K.; Sparkes, R.  
Proc. Natl. Acad. Sci. U.S.A. 86, 8462-8466, 1989  
A:Title: Male-enhanced antigen gene is phylogenetically conserved and expressed at late  
A:Reference number: A34421; MUID:90046817; PMID:2813404  
A:Accession: B34421  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-174 <LAU>  
A:CROSS-references: UNIPROT:Q64327; UNIPARC:UPI0000001930; GB:M27938; NID:gl87507; PIDN:  
A:Note: the authors translated the codon GCA for residue 31 as Ser, and ACA for residue

Query Match 77.8%; Score 35; DB 2; Length 174;  
Best Local Similarity 66.7%; Pred. No. 8.7;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|:|:|:|:|  
Db 154 AQWEDVQK 162

RESULT 4  
A34421  
male-enhanced antigen - human  
C:Species: Homo sapiens (man)  
C:Date: 15-Jun-1990 #sequence\_revision 15-Jun-1990 #text\_change 09-Jul-2004  
C:Accession: A34421  
R:Laue, Y.F.C.; Chan, K.; Sparkes, R.  
Proc. Natl. Acad. Sci. U.S.A. 86, 8462-8466, 1989  
A:Title: Male-enhanced antigen gene is phylogenetically conserved and expressed at late  
A:Reference number: A34421; MUID:90046817; PMID:2813404  
A:Accession: A34421  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-185 <LAU>  
A:CROSS-references: UNIPROT:Q9BV01; UNIPARC:UPI000007291B; GB:M27937; NID:gl87507; PIDN:

Query Match 77.8%; Score 35; DB 2; Length 185;  
Best Local Similarity 66.7%; Pred. No. 9.3;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|:|:|:|:|  
Db 165 AQWEDVQK 173

RESULT 5  
A96941  
hypothetical protein CAC0335 [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: A96941  
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: A96941  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-401 <KUR>  
A:CROSS-references: UNIPROT:Q97M63; UNIPARC:UPI00000C9E2B; GB:AE001437; PIDN:AAK78316.1;

A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0335

Query Match 77.8%; Score 35; DB 2; Length 401;  
Best Local Similarity 66.7%; Pred. No. 22;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|:|:|:|:|  
Db 290 AEFEDAIEK 298

RESULT 6  
G87789  
protein C34G6.7 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: G87789  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_eleg  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: G87789  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-449 <STO>  
A:CROSS-references: UNIPROT:O01498; UNIPARC:UPI000017A544; GB:chr\_I; PIDN:AABS2480.1; PI  
A:Note: contains similarity to an SH3 domain  
C:Genetics:  
A:Gene: C34G6.7  
A:Map position: 1

Query Match 77.8%; Score 35; DB 2; Length 449;  
Best Local Similarity 77.8%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|:|:|:|:|  
Db 357 ALYDDAIQK 365

RESULT 7  
AH1228  
MutS protein (MutS2) homolog lmo1232 [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 31-Dec-2004  
C:Accession: AH1228  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;  
.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant,  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AH1228  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-785 <GLA>  
A:CROSS-references: UNIPROT:Q8Y7P1; UNIPARC:UPI000005509F; GB:NC\_003210; PIDN:CAC99310.1;  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo1232  
C:Superfamily: DNA mismatch repair protein, bacterial/plant types

Query Match 77.8%; Score 35; DB 2; Length 785;  
Best Local Similarity 75.0%; Pred. No. 46;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9

Db 716 RYEDALQK 723  
:||||:|

## RESULT 8

S70115  
Z1P1 protein - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein D9819.9; protein YDR285w  
C;Species: Saccharomyces cerevisiae  
C;Date: 24-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 09-Jul-2004  
C;Accession: S70115; A45173; S30868  
R;Fulton, L.  
submitted to the EMBL Data Library, May 1996  
A;Description: The sequence of *S. cerevisiae* cosmid 9819.  
A;Reference number: S70115  
A;Accession: S70115  
A;Molecule type: DNA  
A;Residues: 1-875 <FUL>  
A;Cross-references: UNIPROT:P31111; UNIPARC:UPI0000013C3E2; EMBL:U51031; NID:gl332635; PIDN:AAAB6710.1  
R;Sym: M.; Engbrecht, J.A.; Roeder, G.S.  
Cell 72, 365-378, 1993  
A;Title: Z1P1 is a synaptonemal complex protein required for meiotic chromosome synapsis  
A;Reference number: A45173; MUID:93161412; PMID:7916652  
A;Accession: A45173  
A;Molecule type: DNA  
A;Residues: 1-54, 'A', 56-875 <SYM>  
A;Cross-references: UNIPARC:UPI0000168FB9; EMBL:L06487; NID:gl732240; PIDN:AAA35239.1; PIDN:AAAB6710.1  
A;Note: sequence extracted from NCBI backbone (NCBIP:124844)  
C;Genetics:  
A;Gene: SGD:Z1P1  
A;Cross-references: SGD:S0002693; MIPS:YDR285w  
A;Map position: 4R

Query Match 77.8%; Score 35; DB 2; Length 875;  
Best Local Similarity 66.7%; Pred. No. 51;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|||||:|  
Db 529 AQYEDLVKK 537

## RESULT 9

T31733  
hypothetical protein T05C3.2 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 03-Nov-2003  
C;Accession: T31733  
R;Blanchard, M.; Bradshaw, H.; Kramer, J.  
submitted to the EMBL Data Library, July 1997  
A;Description: The sequence of *C. elegans* cosmid T05C3.  
A;Reference number: Z21076  
A;Accession: T31733  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2327 <BLA>  
A;Cross-references: UNIPARC:UPI000017BB60; EMBL:AF016428; PIDN:AAAB65359.1; GSPDB:GN00023  
A;Experimental source: strain Bristol N2; clone T05C3  
C;Genetics:  
A;Gene: CESP:T05C3.2  
A;Map position: 5  
A;Introns: 4/1; 45/2; 77/1; 130/3; 184/2; 283/3; 388/1; 431/1; 492/1; 549/1; 645/3; 678/1771/1; 1835/3; 1897/3; 2008/2; 2094/3; 2150/3; 2198/3; 2238/1; 2325/2  
C;Superfamily: phospholipase D, chodopox type

Query Match 77.8%; Score 35; DB 2; Length 2327;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 8  
|||||  
Db 1475 ADYEDAIQ 1482

## RESULT 10

T37196  
hypothetical protein F49D11.4 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37196  
R;Hinds, K.; Graves, T.; Duckels, G.; Ozersky, P.  
submitted to the EMBL Data Library, December 1997  
A;Description: The sequence of *C. elegans* cosmid F49D11.  
A;Reference number: Z21632  
A;Accession: T37196  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-416 <HIN>  
A;Cross-references: UNIPROT:O44723; UNIPARC:UPI000005C8F6; EMBL:AF039711; PIDN:AAAB6710.1  
C;Genetics:  
A;Map position: 1  
A;Introns: 189/1; 282/3; 372/1  
A;Note: F49D11.4

Query Match 75.6%; Score 34; DB 2; Length 416;  
Best Local Similarity 66.7%; Pred. No. 37;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|||||:|  
Db 349 ANYEALQK 357

## RESULT 11

T28386  
hypothetical protein 225 - *Melanoplus sanguinipes* entomopoxvirus  
C;Species: *Melanoplus sanguinipes* entomopoxvirus  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T28386  
R;Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A;Title: The genome of *Melanoplus sanguinipes* entomopoxvirus.  
A;Reference number: Z20484; MUID:99102612; PMID:9847359  
A;Accession: T28386  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-142 <AFO>  
A;Cross-references: UNIPROT:Q9YVL7; UNIPARC:UPI00000F658C; EMBL:AF063866; NID:g4049647; PIDN:AAAB6710.1  
C;Genetics:  
A;Note: MSV225  
C;Superfamily: *Melanoplus sanguinipes* entomopoxvirus hypothetical protein 225

Query Match 73.3%; Score 33; DB 2; Length 142;  
Best Local Similarity 85.7%; Pred. No. 18;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8  
|||||  
Db 84 QYEDAIQ 90

## RESULT 12

T27172  
hypothetical protein Y54G11A.8 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T27172  
R;Wallis, J.  
submitted to the EMBL Data Library, December 1998  
A;Reference number: Z20322  
A;Accession: T27172  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-396 <WIL>  
A;Cross-references: UNIPROT:Q9XW00; UNIPARC:UPI00000753C5; EMBL:AL034488; NID:e1359895; PIDN:AAAB6710.1  
A;Experimental source: clone Y54G11A

C:Genetics:  
A:Gene: CESP.Y54GL1A.8  
A:Introns: 55/2; 99/3; 219/2; 278/3; 307/2; 349/3

Query Match 73.3%; Score 33; DB 2; Length 396;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8  
Db 108 QYDDAIQ 114  
||:||||  
||:||||

RESULT 13  
A49937  
50k outer membrane protein oprK - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C:Date: 11-Nov-1994 #sequence\_revision 11-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A49937; S39631  
R:Poole, K.; Krebs, K.; McNally, C.; Neshat, S.  
J. Bacteriol. 175, 7363-7372, 1993  
A:Title: Multiple antibiotic resistance in Pseudomonas aeruginosa: evidence for involvement of OprK  
A:Reference number: A49937; MUID:94042912; PMID:8226684  
A:Accession: A49937  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-477 <PO>  
A:Cross-references: UNIPROT:Q51487; UNIPARC:UPI000016FCCF; GB:L23839; NID:9404772; PIDN: Mol. Microbiol. 10, 529-544, 1993  
R:Poole, K.; Heinrichs, D.E.; Neshat, S.  
A:Title: Cloning and sequence analysis of an EnvCD homologue in Pseudomonas aeruginosa: A:Reference number: S39629; MUID:95058196; PMID:7968531  
A:Accession: S39631  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-296 <P02>  
A:Cross-references: UNIPARC:UPI000016FC9B; GB:L11616; NID:9438852; PIDN:AAA74438.1; PID: C:Genetics:  
A:Gene: oprK  
C:Superfamily: nodulation protein nodT

Query Match 73.3%; Score 33; DB 2; Length 477;  
Best Local Similarity 87.5%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEAIQ 8  
Db 373 AQYEAIQ 380  
|||||  
|||||

RESULT 14  
F83593  
outer membrane protein OprM precursor PA0427 [imported] - Pseudomonas aeruginosa (strain C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: F83593  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mitzoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B. adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam, .; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: F83593  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-485 <STO>  
A:Cross-references: UNIPROT:Q51487; UNIPARC:UPI0000130D9F; GB:AE004479; GB:AE004091; NID: A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: oprM; PA0427  
C:Superfamily: nodulation protein nodT

Query Match 73.3%; Score 33; DB 2; Length 485;



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 92.0971 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-4

Perfect score: 45

Sequence: 1 AQYEDAIQK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	82.2	785	1 MUTS2_LISIN	Q92ch6 listeria in
2	36	80.0	484	2 Q9F1E5_ARATH	Q9f1e5 arabidopsis
3	36	80.0	559	2 Q8GZW0_ARATH	Q8gzw0 arabidopsis
4	36	80.0	681	1 AC0XZ2_RAT	P97562 r acyl-coen
5	36	80.0	788	2 Q5E2X9_VIBF1	Q5e2x9 vibrio fusc
6	35	77.8	131	2 Q9S387_LISMO	Q9s387 listeria mo
7	35	77.8	172	1 MEAL_HUMAN	Q16626 homo sapien
8	35	77.8	173	2 Q4R952_MACFA	Q4r952 macaca fasc
9	35	77.8	174	1 MEAL_BOVIN	Q29407 bos taurus
10	35	77.8	174	1 MEAL_MOUSE	Q64327 mus musculu
11	35	77.8	174	2 Q3T097_BOVIN	Q3t097 bos taurus
12	35	77.8	174	2 Q5FVH7_RAT	Q5fvh7 rattus norv
13	35	77.8	174	2 Q793G4_MOUSE	Q793g4 m male-enha
14	35	77.8	174	2 Q9DQO0_MOUSE	Q9dqo0 mus musculu
15	35	77.8	174	2 Q9DB93_MOUSE	Q9db93 mus musculu
16	35	77.8	258	2 Q8SW28_ENCCU	Q8sw28 encephalito
17	35	77.8	378	1 CAAL_BUEBA	Q6mkw4 bdellovibri
18	35	77.8	397	2 Q8I7I2_CABEL	Q8i7i2 caenorhabdi
19	35	77.8	401	2 Q97M63_CLOAB	Q97m63 clostridium
20	35	77.8	457	2 Q01498_CABEL	Q01498 caenorhabdi
21	35	77.8	594	2 Q3UV17_MOUSE	Q3uv17 mus musculu
22	35	77.8	607	2 Q6BIU4_DEBHA	Q6biu4 debaryomyce
23	35	77.8	678	2 Q90WB4_BRARE	Q90wb4 brachydanio
24	35	77.8	681	2 Q90X2_MOUSE	Q90xd1 m acyl-coen
25	35	77.8	681	2 Q3UNF3_MOUSE	Q3unf3 mus musculu
26	35	77.8	681	2 Q8VCB0_MOUSE	Q8vcb0 mus musculu
27	35	77.8	752	2 Q6LXF6_METMP	Q6lxf6 methanococc
28	35	77.8	762	2 Q90XS6_BRARE	Q90xs6 brachydanio
29	35	77.8	768	2 Q2T154_PECOC	Q2t154 pectobacter
30	35	77.8	768	2 Q6D3F4_ERWCT	Q6d3f4 erwinia car
31	35	77.8	785	1 MUTS2_LISMF	Q720j7 listeria mo

32	35	77.8	785	1 MUTS2_LISMO	Q8y7p1 listeria mo
33	35	77.8	785	2 Q4EEU4_LISMO	Q4eeu4 listeria mo
34	35	77.8	785	2 Q4ENE4_LISMO	Q4ene4 listeria mo
35	35	77.8	875	1 ZIPI1_YEAST	P31111 saccharomyc
36	35	77.8	893	2 Q6NUY7_BRARE	Q6nuy7 brachydanio
37	35	77.8	893	2 Q7SZR3_BRARE	Q7szr3 brachydanio
38	35	77.8	918	2 Q54DV3_DICDI	Q54dv3 dictyosteli
39	35	77.8	1392	2 Q3VFA6_9SPHN	Q3vfa6 sphingopyxi
40	35	77.8	1729	2 Q61KG0_CABER	Q61kg0 caenorhabdi
41	35	77.8	1735	2 Q16302_CABEL	Q16302 caenorhabdi
42	35	77.8	4421	2 Q4RK60_TETNG	Q4rk60 tetraodon n
43	34	75.6	101	2 Q8DBR4_VIBVU	Q8dbr4 vibrio vuln
44	34	75.6	124	2 Q5NGF8_FRATT	Q5ngf8 franciella
45	34	75.6	155	2 Q7VP20_HAEDU	Q7vp20 haemophilus

#### ALIGNMENTS

#### RESULT 1

ID	MUTS2_LISIN	STANDARD;	PRT;	785 AA.
AC	Q92CH6;			
DT	21-MAY-2002,	integrated into UniProtKB/Swiss-Prot.		
DT	01-DEC-2001,	sequence version 1.		
DT	07-MAR-2006,	entry version 23.		
DE	MutS2 protein.			
GN	Name=mutS2; OrderedLocusNames=lin1195;			
OS	Listeria innocua.			
OC	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.			
OX	NCBI_TaxID=1642;			
RP	[1]			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	STRAIN=CLIP 11262 / Serovar 6a;			
RX	MEDLINE=21537279; PubMed=11679669; DOI=10.1126/science.1063447;			
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,			
RA	Baquerio F., Berche P., Bloeker H., Brandt P., Chakraborty T.,			
RA	Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,			
RA	Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,			
RA	Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,			
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,			
RA	Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,			
RA	Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,			
RA	Nordieck G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,			
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,			
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;			
RT	"Comparative genomics of Listeria species.";			
RL	Science 294:849-852(2001).			
CC	!- FUNCTION: Not known.			
CC	!- SIMILARITY: Belongs to the DNA mismatch repair mutS family.			
CC	!- SIMILARITY: Contains 1 Smr domain.			
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>			
CC	Distributed under the Creative Commons Attribution-NoDerivs License			
DR	EMBL; AL596167; CAC96426.1; -; Genomic_DNA.			
DR	PIR; AB1582; AB1582.			
DR	HSSP; Q56215; LEWR.			
DR	GenomeReviews; AL592022_GR; lin1195.			
DR	ListiList; LIN01195; -.			
DR	BioCyc; LINN1642:LIN1195-MONOMER; -.			
DR	HAMAP; MF_00092; -; 1.			
DR	InterPro; IPR005747; MutS2.			
DR	InterPro; IPR000432; MutS C.			
DR	InterPro; IPR007696; MutS III.			
DR	InterPro; IPR002625; Smr/MutS2_C.			
DR	Pfam; PF00488; MutS_V; 1.			
DR	Pfam; PF01713; Smr; 1.			
DR	PIRSF; PIRSF005814; MutS_yshd; 1.			
DR	ProDom; PD001263; MutS_C; 1.			
DR	SMART; SM00534; MUTSAC; 1.			
DR	SMART; SM00533; MUTSD; 1.			
DR	SMART; SM00463; SMR; 1.			

DR TIGRFAMS; TIGR01069; mutS2; 1.  
 DR PROSITE; PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
 DR PROSITE; PSS0828; SMR; 1.  
 KW ATP-binding; Complete proteome; DNA-binding; Nucleotide-binding.  
 FT CHAIN 1 785 MutS2 protein.  
 FT FT 710 785 Smr.  
 FT NP BIND 335 342 ATP (Potential).  
 SQ SEQUENCE 785 AA; 86679 MW; E5548422PB87CA17 CRC64;

Query Match 82.2%; Score 37; DB 1; Length 785;  
 Best Local Similarity 87.5%; Pred. NO. 1.9e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9  
 DB 716 RYEDAIQK 723

## RESULT 2

Q3FIL5 ARATH PRELIMINARY; PRT; 484 AA.  
 AC Q3FIL5;  
 DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.  
 DT 07-FEB-2006, entry version 12.  
 DE Gb|AAB2637.1 (AT5G58960/k19m22\_160).  
 GN OrderedLocustNames=AT5G58960;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=99156233; PubMed=10048488; DOI=10.1093/dnares/5.6.379;  
 RA Asanizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RA "Structural analysis of Arabidopsis thaliana chromosome 5, VIII.  
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:379-391(1998).  
 RN [2]

RP NUCLEOTIDE SEQUENCE.  
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Banh J., Bowser L.,  
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jones I., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP NUCLEOTIDE SEQUENCE.  
 RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,  
 RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----  
 DR EMBL; AB016885; BAB09639.1; -; Genomic DNA.  
 DR EMBL; AY075634; AAL91642.1; -; mRNA.  
 DR EMBL; BT002623; AAO11539.1; -; mRNA.

DR TAIR; AT5G58960; -;  
 DR InterPro; IPR006943; DUF641\_pln.  
 DR Pfam; PF04859; DUF641; 1.

SEQUENCE 484 AA; 54071 MW; FDCEFE146D40782F CRC64;

Query Match 80.0%; Score 36; DB 2; Length 484;  
 Best Local Similarity 75.0%; Pred. NO. 1.9e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9  
 DB 53 QYEDALEK 60

## RESULT 3

Q8GWZ0 ARATH PRELIMINARY; PRT; 559 AA.  
 AC Q8GWZ0;  
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 8.  
 DE Hypothetical protein At5G58960/k19m22\_160.  
 GN Name=At5G58960/k19m22\_160;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; AK118548; BAC43150.1; -; mRNA.  
 DR InterPro; IPR006943; DUF641\_pln.  
 DR Pfam; PF04859; DUF641; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 559 AA; 62668 MW; 3C0FCAE17FD0F801 CRC64;

Query Match 80.0%; Score 36; DB 2; Length 559;  
 Best Local Similarity 75.0%; Pred. NO. 2.2e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9  
 DB 128 QYEDALEK 135

## RESULT 4

ACOX2 RAT STANDARD; PRT; 681 AA.  
 AC P97562;  
 DT 02-NOV-2001, integrated into UniProtKB/Swiss-Prot.  
 DT 01-MAY-1997, sequence version 1.  
 DT 07-MAR-2006, entry version 41.  
 DE Acyl-coenzyme A oxidase 2, peroxisomal (EC 1.17.99.3) (3-alpha,7-alpha,12-alpha-trihydroxy-5-beta-cholestanoyl-CoA 24-hydroxylase) (3-alpha,7-alpha,12-alpha-trihydroxy-5-beta-cholestanoyl-CoA oxidase).  
 DE (Trihydroxycoprostanoyl-CoA oxidase) (THCCox) (THCA-CoA oxidase).  
 GN Name=Acox2; Synonyms=Thcox;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridea; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]

RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 27-33; 74-83; 128-137;  
 RP 265-270; 328-336; 454-464 AND 656-667, AND FUNCTION.

RC TISSUE=Liver;  
 RX MEDLINE=97103103; PubMed=8947475;

RA Baumgart E., Vanhooren J.C.T., Franssen M., Van Leuven F., Fahimi H.D.,  
 RA Van Veldhoven P.P., Mannaerts G.P.;

RT "Molecular cloning and further characterization of rat peroxisomal  
 RT trihydroxycoprostanoyl-CoA oxidase.";  
 RL Biochem. J. 320:115-121 (1996).  
 CC -1- FUNCTION: Oxidizes the CoA esters of the bile acid intermediates  
 CC di- and tri-hydroxycoprostanic acids.  
 CC -1- CATALYTIC ACTIVITY: (2S)-3-alpha,7-alpha,12-alpha-trihydroxy-5-  
 CC beta-cholestan-26-oyl-CoA + H(2)O + acceptor = (24R,25R)-3-  
 CC alpha,7-alpha,12-alpha,24-tetrahydroxy-5-beta-cholestan-26-oyl-CoA  
 CC + reduced acceptor.  
 CC -1- COFACTOR: FAD (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Peroxisome.  
 CC -1- TISSUE SPECIFICITY: Most abundant in liver. Also expressed in  
 CC kidney. Not present in any other tissues tested.  
 CC -1- SIMILARITY: Belongs to the acyl-CoA oxidase family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; X95189; CAA64488.1; -; mRNA.  
 DR HSSP; P07872; IIS2.  
 DR EMBL; ENSRNOG00000007378; Rattus norvegicus.  
 DR Ensembl; ENSRNOG00000007378; Rattus norvegicus.  
 DR RGD; 628684; Acox2.  
 DR GO; GO:0005777; C:peroxisome; IDA.  
 DR GO; GO:0006631; P:fatty acid metabolism; IEP.  
 DR InterPro; IPR012258; Acyl-CoA oxidase.  
 DR InterPro; IPR06091; Acyl-CoA DH M.  
 DR InterPro; IPR02655; Acyl-CoA ox C.  
 DR Pfam; PF01756; ACOX; 1.  
 DR Pfam; PF02770; Acyl-CoA oxidase; 1.  
 DR PRSF; PRSF000168; Acyl-CoA oxidase; 1.  
 KW Direct protein sequencing; FAD; Fatty acid metabolism; Flavoprotein;  
 KW Lipid metabolism; Oxidoreductase; Peroxisome.  
 FT CHAIN 1 681  
 FT Acyl-coenzyme A oxidase 2.  
 FT /FTID:PRO 0000204684.  
 FT MOTIF 679 681 Microbody targeting signal (Potential).  
 SQ SEQUENCE 681 AA; 76799 MW; 3599D71365D38932 CRC64;  
 Query Match 80.0%; Score 36; DB 1; Length 681;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 YEDAIOK 9  
 Db 77 YEDAIOK 83  
 RESULT 5  
 QSE2X9\_VIBF1 PRELIMINARY; PRT; 788 AA.  
 AC QSE2X9;  
 DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.  
 DT 15-MAR-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE Aerobic respiration control sensor protein ArcB (EC 2.7.3.-).  
 GN OrderedLocustNames=VF2122;  
 OS Vibrio fischeri (strain ATCC 700601 / ES114).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 OX NCBI\_TaxID=312309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RX PubMed=15703294; DOI=10.1073/pnas.0409900102;  
 RA Ruby E.G., Urbanowski M., Campbell J., Dunn A., Faini M., Gunsalus R.,  
 RA Lostro P., Lupp C., McCann J., Millikan D., Schaefer A., Stabb E.,  
 RA Stevens A., Visick K., Whistler C., Greenberg E.P.;  
 RT "Complete genome sequence of Vibrio fischeri: a symbiotic bacterium  
 RT with pathogenic congeners.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 102:3004-3009 (2005).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; CP000020; AAW86617.1; -; Genomic\_DNA.

DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0016301; F:kinase activity; IEA.  
 DR GO; GO:0000156; F:two-component response regulator activity; IEA.  
 DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR GO; GO:0000160; P:two-component signal transduction system (p...); IEA.  
 DR InterPro; IPR003594; ATP bd ATPase.  
 DR InterPro; IPR004358; His\_kin\_like\_C.  
 DR InterPro; IPR003661; His\_kin\_N.  
 DR InterPro; IPR005467; His\_kinase.  
 DR InterPro; IPR008207; Hpt.  
 DR InterPro; IPR000014; PAS.  
 DR InterPro; IPR001789; Response\_reg.  
 DR Pfam; PF02518; HATPase\_C; 1.  
 DR Pfam; PF00512; HSKA; 1.  
 DR Pfam; PF01627; Hpt; 1.  
 DR Pfam; PF00989; PAS; 1.  
 DR Pfam; PF00072; Response\_reg; 1.  
 DR PRINTS; PR00344; BCTRLSENSOR.  
 DR ProDom; PD000039; Response\_reg; 1.  
 DR SMART; SM00387; HATPase\_C; 1.  
 DR SMART; SM00388; HSKA; 1.  
 DR SMART; SM00091; PAS; 1.  
 DR SMART; SM00091; PAS; 1.  
 DR SMART; SM00448; REC; 1.  
 DR TIGRFAMS; TIGR00229; sensory\_box; 1.  
 DR PROSITE; PS50109; HIS\_KIN; 1.  
 DR PROSITE; PS50894; HPT; 1.  
 DR PROSITE; PS50113; PAC; 1.  
 DR PROSITE; PS50112; PAS; 1.  
 DR PROSITE; PS50110; RESPONSE\_REGULATORY; 1.  
 KW Complete proteome; Transferase.  
 SQ SEQUENCE 788 AA; 89466 MW; 0BA4DF7E54A4FAF7 CRC64;  
 Query Match 80.0%; Score 36; DB 2; Length 788;  
 Best Local Similarity 75.0%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QYEDAIQK 9  
 Db 271 QYEDAIQK 278  
 RESULT 6  
 Q9S387\_LISMO PRELIMINARY; PRT; 131 AA.  
 ID Q9S387\_LISMO PRELIMINARY; PRT; 131 AA.  
 AC Q9S387;  
 DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
 DT 01-MAY-2000, sequence version 1.  
 DT 07-FEB-2006, entry version 16.  
 DE MutS-like protein (Fragment).  
 GN Name=mutS;  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=EGD;  
 RX MEDLINE=98196733; PubMed=9537387;  
 RA Scharf C., Riechdorf S., Ernst H., Engelmann S., Volker U., Hecker M.;  
 RT "Thioredoxin is an essential protein induced by multiple stresses in  
 RT Bacillus subtilis.";  
 RL J. Bacteriol. 180:1869-1877 (1998).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=EGD;  
 RA Tarchanov M., Borovok I., Aharonowitz Y., Cohen G.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=EGD;
RA Borovok I., Mislovat M., Cohen G., Aharonowitz Y.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=EGD;
RA Borovok I., Aharonowitz Y., Cohen G.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=EGD;
RA Borovok I.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ133006; CAB40814.2; -; Genomic_DNA.
DR InterPro; IPR002625; Smr/MutS2_C.
DR Pfam; PF01713; Smr; 1.
DR SMART; SM00463; Smr; 1.
DR PROSITE; PS50828; Smr; 1.
FT NON TER 1
SQ SEQUENCE 131 AA; 14481 MW; 7C401C6035798066 CRC64;

Query Match 77.8%; Score 35; DB 2; Length 131;
Best Local Similarity 75.0%; Pred. No. 73;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 QVEDAIQK 9
DB 62 RYEDALQK 69

RESULT 7
MEAL HUMAN
ID MEAL HUMAN STANDARD; PRT; 172 AA.
AC Q16626; Q5TC36; Q9BV01;
DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 1.
DT 07-FEB-2006, entry version 34.
DE Male-enhanced antigen 1 (Mea-1).
GN Name=MEAL; Synonyms=MEA;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC TISSUE=Testis;
RX MEDLINE=90046817; PubMed=2813404;
RA Lau Y.-F.C., Chan K., Sparkes R.S.;
RT "Male-enhanced antigen gene is phylogenetically conserved and
RT expressed at late stages of spermatogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8462-8466(1989).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RA Kalnine N., Chen X., Roifs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length CDSs in BD Creator(TM) system donor
RT vector.";
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=22935763; PubMed=14574404; DOI=10.1038/nature02055;
RA Mungall A.J., Palmer S.A., Sims S.K., Edwards C.A., Ashurst J.L.,
RA Wilming L., Jones M.C., Horton R., Hunt S.E., Scott C.E.,
RA Gilbert J.G.R., Clamp M.E., Bethel G., Milne S., Ainscough R.,
RA Almeida J.P., Ambrose K.D., Andrews T.D., Ashwell R.I.S.,
RA Babbage A.K., Bagguley C.L., Bailey J., Banerjee R., Barker D.J.,

```

```

RA Barlow K.F., Bates K., Beare D.M., Beasley H., Beasley O., Bird C.F.,
RA Blakey S.E., Bray-Allen S., Brook J., Brown A.J., Brown J.Y.,
RA Burford D.C., Burrill W., Burton J., Carder C., Carter N.P.,
RA Chapman J.C., Clark S.Y., Clark G., Clee C.M., Clegg S., Cobley V.,
RA Collier R.E., Collins J.E., Colman L.K., Corby N.R., Coville G.J.,
RA Culley K.M., Dhani P., Davies J., Dunn M., Earthrowl M.E.,
RA Ellington A.E., Evans K.A., Faulkner L., Francis M.D., Frankish A.,
RA Frankland J., French L., Garner P., Garnett J., Ghori M.J.,
RA Gibby L.M., Gillson C.J., Glithero R.J., Grafham D.V., Grant M.,
RA Gribble S., Griffiths C., Griffiths M.N.D., Hall R., Halls K.S.,
RA Hammond S., Harley J.L., Hart E.A., Heath P.D., Heathcote R.,
RA Holmes S.J., Howden P.J., Howe K.L., Howell G.R., Huckle E.,
RA Humphray S.J., Humphries M.D., Hunt A.R., Johnson C.M., Joy A.A.,
RA Kay M., Keenan S.J., Kimberley A.M., King A., Laird G.K., Langford C.,
RA Lawlor S., Leongamornlert D.A., Leversha M., Lloyd C.R., Lloyd D.M.,
RA Loveland J.E., Lovell J., Martin S., Mashreghi-Mohammadi M.,
RA Mathew G.B., Mathews L., McCann O.T., McLaren S.J., McLeay K.,
RA McMurray A., Moore M.J.F., Mullikin J.C., Niblett D., Nickerson T.,
RA Novik K.L., Oliver K., Overton-Larty E.K., Parker A., Patel R.,
RA Pearce A.V., Peck A.I., Phillimore B.J.C.T., Phillips S., Plumb R.M.,
RA Porter K.M., Ramsey Y., Ranby S.A., Rice C.M., Ross M.T., Searle S.M.,
RA Sehra H.K., Sheridan E., Skuce C.D., Smith S., Smith M., Spraggon L.,
RA Squares S.L., Stewart C.A., Sycamore N., Tamlyn-Hall G., Tester J.,
RA Theaker A.J., Thomas D.W., Thorpe A., Tracey A., Tromans A., Tubby B.,
RA Whittaker H., Wild A., Willey D.J., Wilmer T.E., Wood J.M., Wray P.W.,
RA Wyatt J.C., Young L., Younger R.M., Bentley D.R., Coulson A.,
RA Durbin R., Hubbard T., Sulston J.E., Dunham I., Rogers J., Beck S.;
RT "The DNA sequence and analysis of human chromosome 6.";
RN Nature 425:805-811(2003).
[4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Eye;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play an important role in spermatogenesis and/or
CC testis development.
CC -!- INTERACTION:
CC P63010; AP2B1; NbExp=1; IntAct=EBI-744921; EBI-432924;
CC -!- TISSUE SPECIFICITY: Highly expressed in testis.
CC -!- CAUTION: Was originally thought to be the H-Y antigen.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-11 is the initiator.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; M27937; AAA36208.1; ALT_INIT; mRNA.
DR EMBL; L10400; AAA36209.1; ALT_INIT; mRNA.
DR EMBL; BT009831; AAP88833.1; ALT_INIT; mRNA.
DR EMBL; AL136304; CA119794.1; -; Genomic DNA.
DR EMBL; BC001754; AA01754.1; ALT_INIT; mRNA.
DR PIR; A34421; A34421.
DR IntAct; Q16626; -.
DR Ensembl; ENSG00000124733; Homo sapiens.

```

```
DR H-InvDB; HIX0005890; -.
DR HGNC; HGNC:6986; MEAL.
DR MIM; 143170; Gene.
DR LinkHub; Q16626; -.
DR GO; GO:0008584; P:male gonad development; TAS.
DR GO; GO:0007283; P:spermatogenesis; TAS.
DR InterPro; IPR009685; MEAL.
DR PANTHER; PTHR17005; MEAL; 1.
DR Pfam; PF06910; MEAL; 1.
KW Developmental protein; Differentiation; Spermatogenesis.
FT CHAIN 1 172 Male-enhanced antigen 1.
FT /FTID=PRO 0000096341.
SQ SEQUENCE 172 AA; 18544 MW; 9342AFAD2C56FF25 CRC64;

Query Match 77.8%; Score 35; DB 1; Length 172;
Best Local Similarity 66.7%; Pred. No. 98;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9
Db 152 AQWEDVVQK 160

RESULT 8
Q4R992_MACFA PRELIMINARY; PRT; 173 AA.
ID Q4R992;
AC Q4R992;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Testis cDNA clone: Q5A-10511, similar to human male-enhanced antigen (MEAL).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
OC Cercopithecoidea; Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15944441; DOI=10.1093/molbev/msl187;
RA Osada N., Hirata M., Tanuma R., Kusuda J., Hida M., Suzuki Y.,
RA Sugano S., Gojobori T., Shen C.-K.J., Wu C.I., Hashimoto K.;
RT "Substitution Rate and Structural Divergence of 5'UTR Evolution;
RT Comparative Analysis Between Human and Cynomolgus Monkey cDNAs.";
RL Mol. Biol. Evol. 22:1976-1982(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG International consortium for macaque cDNA sequencing and analysis;
RT "DNA sequences of macaque genes expressed in brain or testis and its
RT evolutionary implications.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AB168204; BAE00329.1; -; mRNA.
DR GO; GO:0007283; P:spermatogenesis; IEA.
DR InterPro; IPR009685; MEAL.
DR PANTHER; PTHR17005; MEAL; 1.
DR Pfam; PF06910; MEAL; 1.
SQ SEQUENCE 173 AA; 18641 MW; 8B048AB82C0F4E3E CRC64;

Query Match 77.8%; Score 35; DB 2; Length 173;
Best Local Similarity 66.7%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9
Db 153 AQWEDVVQK 161

RESULT 9
MEAL_BOVIN
```

```
ID MEAL_BOVIN STANDARD; PRT; 174 AA.
AC Q29407;
DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 1.
DT 07-FEB-2006, entry version 29.
DE Male-enhanced antigen 1 (Mea-1).
GN Name=MEAL; Synonyms=MEA;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]_TaxID=9913;
RP NUCLEOTIDE SEQUENCE [MRNA].
RC TISSUE=Testis;
RX MEDLINE=94092744; PubMed=8268232; DOI=10.1016/0167-4781(93)90019-A;
RA Kondo M., Sato S., Suto S.;
RT "Cloning and sequence analysis of cDNA encoding the bovine testis-
RT derived male-enhanced antigen (Mea).";
RL Biochim. Biophys. Acta 1216:483-486(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX MEDLINE=97063437; PubMed=8907304;
RA Kondo M., Terouchi S., Tsukasa N., Sato S., Ishida N., Suto S.;
RT "Genomic sequence analysis of the bovine male-enhanced antigen-1 (Mea-
RT 1) and differential localization of its transcripts and products
RT during spermatogenesis.";
RL DNA Seq. 6:75-85(1996).
CC -!- FUNCTION: May play an important role in spermatogenesis and/or
CC testis development.
CC -!- TISSUE SPECIFICITY: Highly expressed in testis. Transcripts can be
CC found in primary and secondary spermatocytes, and spermatids, but
CC the protein itself is only detected in spermatids. No expression
CC in Leydig cells, spermatogonia, or sperm. Very weak expression in
CC the heart, kidney, spleen, thymus and ovary.
CC -!- DEVELOPMENTAL STAGE: Expressed in the stages from 8-cell embryos
CC to hatched blastocysts.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-11 is the initiator.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; D17340; BAA04158.1; -; mRNA.
DR EMBL; D30811; BAA06488.1; -; Genomic DNA.
DR PIR; S39790; S39790.
DR InterPro; IPR009685; MEAL.
DR PANTHER; PTHR17005; MEAL; 1.
DR Pfam; PF06910; MEAL; 1.
KW Developmental protein; Differentiation; Spermatogenesis.
FT CHAIN 1 174 Male-enhanced antigen 1.
FT /FTID=PRO 0000096340.
SQ SEQUENCE 174 AA; 18751 MW; E0E7CE95C0CB3530 CRC64;

Query Match 77.8%; Score 35; DB 1; Length 174;
Best Local Similarity 66.7%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9
Db 154 AQWEDVVQK 162

RESULT 10
MEAL_MOUSE STANDARD; PRT; 174 AA.
ID MEAL_MOUSE
AC Q64327;
DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 1.
DT 07-FEB-2006, entry version 33.
DE Male-enhanced antigen 1 (Mea-1).
GN Name=Meal; Synonyms=Mea;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



ID OSFVH7\_RAT PRELIMINARY; PRT; 174 AA.  
 AC OSFVH7;  
 DT 01-MAR-2005, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 7.  
 DE Hypothetical protein.  
 DE Rattus norvegicus (Rat).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=101116;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Ziegler B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RL and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Brain;  
 RA Director MGC Project;  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DDBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; BC089978; AAH89978.1; -; mRNA.  
 DR GO; GO:0007283; P: spermatogenesis; IEA.  
 DR InterPro; IPR009685; MEAL.  
 DR PANTHER; PTHR17005; MEAL; 1.  
 DR Pfam; PF06910; MEAL; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 174 AA; 18584 MW; C3B16361635B176F CRC64;  
 Query Match 77.8%; Score 35; DB 2; Length 174;  
 Best Local Similarity 66.7%; Pred. No. 9;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AQVEDAIQK 9  
 ||:|:|:|  
 Db 154 AQWEDVVQK 162  
 RESULT 13  
 ID Q793G4\_MOUSE PRELIMINARY; PRT; 174 AA.  
 AC Q793G4;  
 DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.  
 DT 10-MAY-2005, sequence version 1.  
 DE Male-enhanced antigen-1 (Adult male cerebellum cDNA, RIKEN full-length  
 DE enriched library, clone:1500038006 product:male enhanced antigen 1,  
 DE full insert sequence) (Adult male small intestine cDNA, RIKEN full-  
 DE length enriched library, clone:2010011M24 product:male enhanced  
 DE antigen 1, full insert sequence) (Adult male small intestine cDNA,

DE RIKEN full-length enriched library, clone:2010103G23 product:male  
 DE enhanced antigen 1, full insert sequence) (13 days embryo liver cDNA,  
 DE RIKEN full-length enriched library, clone:2500003F22 product:MALE-  
 DE ENHANCED ANTIGEN-1 (MEA-1) homolog) (13 days embryo liver cDNA, RIKEN  
 DE full-length enriched library, clone:2510004F20 product:MALE-ENHANCED  
 DE ANTIGEN-1 (MEA-1) homolog) (13 days embryo liver cDNA, RIKEN full-  
 DE length enriched library, clone:2510012G06 product:MALE-ENHANCED  
 DE ANTIGEN-1 (MEA-1) homolog) (11 days embryo whole body cDNA, RIKEN  
 DE full-length enriched library, clone:2700096H08 product:male enhanced  
 DE antigen 1, full insert sequence).  
 GN Name=Meal; Synonyms=meal;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=129SVJ.  
 RX MEDLINE=22330726; PubMed=12444059; DOI=10.1095/biolreprod.101.002550;  
 RA Ohinata Y., Sutou S., Kondo M., Takahashi T., Mitsui Y.;  
 RT "Male-enhanced antigen-1 gene flanked by two overlapping genes is  
 RL expressed in late spermatogenesis.";  
 RL Biol. Reprod. 67:1824-1831(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole  
 RC body;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole  
 RC body;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilmong L.G., Aldin J.E., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Hummel L., Iacono M., Iakeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Nimomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shitaki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,



RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome."; Science 309:1559-1563(2005).  
 RL [4]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole body;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome."; Science 309:1564-1566(2005).  
 RL [5]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole body;  
 RX PubMed=23254683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaide I., Ooto N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltas L., Marchionni L., McKensie L., Miki H., Nagashima T., Nunata K., Okido T., Pavan W.J., Perte G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wymsh-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."; Nature 420:563-573(2002).  
 RL [6]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole body;  
 RX PubMed=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wymsh-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."; Nature 409:685-690(2001).  
 RL [7]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body; DOI=10.1016/S0076-6879(99)03004-9;  
 RX MEDLINE=99279253; PubMed=10349636;

RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole body;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes."; Genome Res. 10:1617-1630(2000).  
 RL [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole body;  
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 RL [9]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Liver, Small intestine, and Whole body;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H., Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M., Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F., Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H., Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T., Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AB074009; BAB91440.1; -; Genomic DNA.

Query Match 77.8%; Score 35; DB 2; Length 174;

Best Local Similarity 66.7%; Pred. No. 99;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9

||:|:|:|

Db 154 AQMEDVVQK 162

RESULT 14

Q9D0Q0 MOUSE

ID Q9D0Q0 MOUSE PRELIMINARY; PRT; 174 AA.

AC Q9D0Q0

DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.

DT 01-JUN-2001, sequence version 1.

DT 07-FEB-2006, entry version 19.

DE 10 days embryo whole body cDNA, RIKEN full-length enriched library,

DE clone:2600013A18 product:hypothetical protein, full insert sequence.

GN Name=Meal;

OS Mus musculus (Mouse);

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Whole body; DOI=10.1016/S0076-6879(99)03004-9;

RX MEDLINE=99279253; PubMed=10349636;



- RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44 (1999).  
RN [2]
- RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
Oyama K., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
Davis M.J., Wilming L.G., Adkins V., Allen J.E.,  
Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
Hill D., Huminek L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
Motteguit-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
Nikolov R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
Petkovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
Sperling S., Stupka E., Sugtara K., Sultana R., Takenaka Y., Taki K.,  
Tamajima K., Tan S.I., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
Ueda H.R., Van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
Grimmond S.M., Tesardate R.D., Liu E.T., Brusic V., Quackenbush J.,  
Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563 (2005).  
RN [3]
- RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
(Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566 (2005).  
RN [4]
- RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
Yagi K., Tomaru Y., Hasegawa Y., Nogi A., Schonbach C., Gojobori T.,  
Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
Nagashima T., Numa K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
Petkovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
- RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
Sultana R., Takenaka Y., Taylor M.S., Tesardate R.D., Tomita M.,  
Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,  
Miyazaki A., Sakai K., Sasaki K., Shibata K., Shingawa A.,  
Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
RN [5]
- RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,  
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690 (2001).  
RN [6]
- RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630 (2000).  
RN [7]
- RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishida T., Harada A.,  
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771 (2000).  
RN [8]
- RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Whole body;  
RX STRAIN=C57BL/6J; TISSUE=Whole body;  
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai K., Sakai K.,  
Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,

RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK011189; BAB27454.1; -; mRNA.  
DR Ensembl; ENSMUSG00000002768; Mus musculus.  
DR MGI; MGI:96955; Mea1.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR InterPro; IPR009685; MEAL1.  
DR PANTHER; PTHR17005; MEAL1.  
DR Pfam; PF06910; MEAL1.  
KW Hypothetical protein.  
SQ SEQUENCE 174 AA; 18556 MW; C3B25352635B176F CRC64;

Query Match 77.8%; Score 35; DB 2; Length 174;  
Best Local Similarity 66.7%; Pred. No. 99;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
||:||||:  
DB 154 AQWEDVQK 162

RESULT 15  
Q9DB93 MOUSE  
ID Q9DB93\_MOUSE PRELIMINARY; PRT; 174 AA.  
AC Q9DB93;  
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2001, sequence version 1.  
DT 07-FEB-2006, entry version 19.  
DE Adult male cerebellum cDNA, RIKEN full-length enriched library,  
DE clone:1500002806 product:male enhanced antigen 1, full insert  
DE sequence.  
GN Name=Meal;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning."  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilmong L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Anweiler K., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Belsel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gusninch S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottacqui-Fabre S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,

RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Sheng Y.,  
RA Schonbach C., Sekiguchi K., Sempile C.A., Seno S., Sesca L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugiyura K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yananishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kal C., Sakai D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome."  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome."  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Belsel K.W.,  
RA Dalka E., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Grimmond S., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reid D.J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner T., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Mateau Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakauchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohata E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AK005106; BAB23822.1; -; mRNA.  
DR Ensembl; ENSMUSG00000002768; Mus musculus.  
DR MGI; MGI:96957; Mea1.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR InterPro; IPR009685; MEA1.  
DR PANTHER; PTHR17005; MEA1; 1.  
DR Pfam; PF06910; MEA1; 1.  
DR SQU SEQUENCE 174 AA; 18597 MW; C3B173BB0F736B9D CRC64;

Query Match 77.8%; Score 35; DB 2; Length 174;  
Best Local Similarity 66.7%; Pred. No. 99;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQYDAIQK 9  
Db 154 AQWEDVVQK 162

Search completed: July 5, 2006, 19:01:27  
Job time : 95.1971 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 78.2039 Seconds  
(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928A-4  
Perfect score: 45  
Sequence: 1 AQVEDAIQK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 8.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*
- 10: Geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	9	6	AAO26384 Psoriasis
2	45	100.0	9	9	ADV04409 Leishmania
3	35	77.8	9	6	AAO26382 Psoriasis
4	35	77.8	9	9	ADV04407 Leishmania
5	35	77.8	164	2	AAR67587 Bovine ma
6	35	77.8	172	7	ADE62272 Human Pro
7	35	77.8	174	2	AAR67586 Murine ma
8	35	77.8	185	6	ABU07424 Protein d
9	35	77.8	185	9	ADX07373 Cyclin-de
10	35	77.8	644	9	AEA15501 Human pol
11	35	77.8	785	5	ABBA48335 Listeria
12	35	77.8	785	6	ABU32679 Protein e
13	35	77.8	875	8	ADN19018 Bacterial
14	34	75.6	9	4	AAM99474 Vaccine r
15	34	75.6	124	8	ADH12886 Francisel
16	34	75.6	233	4	AAU35049 Enterococ
17	34	75.6	233	6	ABU29360 Protein e
18	33	73.3	117	7	ADM04300 Human pro
19	33	73.3	117	9	AEC87230 Human cdn
20	33	73.3	228	4	AAB94159 Human pro
21	33	73.3	396	6	AAO19957 C elegans
22	33	73.3	420	4	ABU52997 Human tes
23	33	73.3	527	7	ABO68208 Pseudomon

24	33	73.3	590	5	AAU74362	Aau74362 Human cyt
25	33	73.3	808	6	ABU24980	Abu24980 Protein e
26	33	73.3	831	7	ADE63406	Ade63406 Rat Prote
27	33	73.3	831	7	ADE58908	Ade58908 Rat Prote
28	33	73.3	831	7	ADD46632	Add46632 Rat Prote
29	33	73.3	808	6	ABU19261	Abu19261 Protein e
30	33	73.3	1072	5	ABBS7023	Abbs7023 Mouse isc
31	33	73.3	1072	9	ADX26395	Adx26395 Novel cel
32	33	73.3	10	10	AEF10781	Aef10781 Human PRL
33	32	71.1	126	4	AAU41376	Aau41376 Propionib
34	32	71.1	126	6	ABM37895	Abm37895 Propionib
35	32	71.1	129	8	ADQ08004	Adq08004 Fusobacte
36	32	71.1	129	8	ADP70200	Adp70200 Fusobacte
37	32	71.1	148	7	ADC99072	Adc99072 Human KPP
38	32	71.1	149	7	ADC99073	Adc99073 Human KPP
39	32	71.1	173	2	AAW37914	Aaw37914 Disease a
40	32	71.1	173	3	AAV92174	Aay92174 Human car
41	32	71.1	173	5	ABB90758	Abb90758 Human Tum
42	32	71.1	173	6	ABU54465	Abu54465 Human Tum
43	32	71.1	173	7	ADE95438	Ade95438 Human PRL
44	32	71.1	173	7	ADE95441	Ade95441 Human PRL
45	32	71.1	173	8	ADI28335	Adi28335 Human pro

#### ALIGNMENTS

RESULT 1  
AAO26384  
ID AAO26384 standard; peptide; 9 AA.  
XX  
AC AAO26384;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 4.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
(AKIV-) AKIVA LLC.  
XX  
O'daly JA;  
XX  
WPI; 2003-018763/01.  
XX  
New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
Claim 1; Page 43; 56pp; English.  
XX  
The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention

SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 6; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
 |||||  
 Db 1 AQYEDAIQK 9

RESULT 2  
 ADV04409  
 ID ADV04409 standard; peptide; 9 AA.  
 XX  
 AC ADV04409;  
 XX  
 DT 24-FEB-2005 (first entry)  
 XX  
 DE Leishmania immunotherapeutic peptide SEQ ID NO:4.  
 XX  
 KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;  
 KW psoriasis.  
 XX  
 OS Leishmania sp.  
 XX  
 PN US2004241168-A1.  
 XX  
 PD 02-DEC-2004.  
 XX  
 PF 09-FEB-2004; 2004US-00774928.  
 XX  
 PR 16-MAR-2001; 2001US-00809003.  
 PR 17-OCT-2003; 2003US-00687892.  
 XX  
 PA (ODAL/) ODALY J A.  
 PI Odaly JA;  
 XX  
 PI WPI; 2005-011563/01.  
 DR  
 XX  
 PT Inhibiting selectively T-cell rolling in human, by administering compound  
 PT interfering with cutaneous lymphocyte antigen-E selectin interaction and  
 PT leukocyte function associated antigen-1 interaction.  
 XX  
 PS Claim 9; SEQ ID NO 4; 21pp; English.  
 XX  
 CC The invention relates to a novel method for selectively inhibiting T-cell  
 CC rolling in a human host, comprising administering a compound that  
 CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
 CC selectin interaction and leukocyte function associated antigen (LFA)-  
 CC 1/intercellular adhesion molecule (ICAM) and very late antigen  
 CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
 CC includes an immunotherapeutic agent, which comprises a purified protein  
 CC extract that is isolated by diethylaminoethyl Sephadex chromatography of  
 CC a Nonidet P-40 insoluble particulate antigen fraction derived from  
 CC isolated killed cells of amastigotes from one or more species of the  
 CC Leishmania genus, where the particulate antigen fraction is solubilized  
 CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
 CC to diethylaminoethyl Sephadex and is eluted with a solution comprising  
 CC 0.1 M sodium chloride, 8 M urea and 0.025 M  
 CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract  
 CC includes polypeptides having apparent molecular weights of 73, 80 and 82  
 CC kDa, after total reduction and alkylation. The species is L. amazonensis,  
 CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the  
 CC invention has antipsoriatic and immunostimulant activity, and inhibits T-  
 CC cell rolling by interfering with CLA-E selectin interaction and LFA-  
 CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
 CC inhibiting T-cell rolling in a human host, and for treating psoriasis.  
 CC The present sequence represents a Leishmania peptide used in the  
 CC invention.  
 XX  
 XX Sequence 9 AA;

Query Match 77.8%; Score 35; DB 6; Length 9;  
 Best Local Similarity 77.8%; Pred. No. 2.1e+06;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
 |||||  
 Db 1 AQYEDAIQK 9

RESULT 4  
 ADV04407  
 ID ADV04407 standard; peptide; 9 AA.  
 XX  
 AC ADV04407;  
 XX

Query Match 100.0%; Score 45; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
 |||||  
 Db 1 AQYEDAIQK 9

RESULT 3  
 AAO26382  
 ID AAO26382 standard; peptide; 9 AA.  
 XX  
 AC AAO26382;  
 XX  
 DT 30-JAN-2003 (first entry)  
 XX  
 DE Psoriasis treating immunotherapeutic peptide, SEQ ID NO 2.  
 XX  
 KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
 KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
 KW gene therapy.  
 XX  
 OS Leishmania sp.  
 XX  
 PN WO200274239-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 04-MAR-2002; 2002WO-US006496.  
 XX  
 PR 16-MAR-2001; 2001US-00809003.  
 XX  
 PA (AKIV-) AKIVA LLC.  
 PI O'daly JA;  
 XX  
 PI WPI; 2003-018763/01.  
 DR  
 XX  
 PT New polypeptide and immunogenic variants comprising amino acid sequences  
 PT of particulate antigens, useful for the treatment and clinical remission  
 PT of psoriasis.  
 XX  
 PS Claim 1; Page 43; 56pp; English.  
 XX  
 CC The invention relates to a polypeptide comprising an isolated amino acid  
 CC sequence or immunogenic variants selected from any of 14 fully defined  
 CC sequences of 7-16 amino acids, given in the specification. The  
 CC immunotherapeutic agents and a pharmaceutical compositions comprising  
 CC polynucleotides and vectors of the invention are useful for the treatment  
 CC and clinical remission of psoriasis. The isolated nucleic acids are  
 CC useful as probes. The sequences of the invention can be used in the  
 CC treatment of disorders by gene therapy. This sequence represents one of  
 CC the 14 immunotherapeutic peptides of the invention  
 XX  
 SQ Sequence 9 AA;

DT 24-FEB-2005 (first entry)  
 XX Leishmania immunotherapeutic peptide SEQ ID NO:2.  
 DE T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;  
 XX psoriasis.  
 KW Leishmania sp.  
 OS US2004241168-A1.  
 XX 02-DEC-2004.  
 PN 09-FEB-2004; 2004US-00774928.  
 XX 16-MAR-2001; 2001US-00809003.  
 PR 17-OCT-2003; 2003US-00687892.  
 XX (ODAL/) ODALY J A.  
 PA Odaly JA;  
 PI WPI; 2005-011563/01.  
 DR Inhibiting selectively T-cell rolling in human, by administering compound  
 XX interfering with cutaneous lymphocyte antigen-E selectin interaction and  
 PT leukocyte function associated antigen-1 interaction.  
 XX Claim 9; SEQ ID NO 2; 21pp; English.  
 PS The invention relates to a novel method for selectively inhibiting T-cell  
 CC rolling in a human host, comprising administering a compound that  
 CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
 CC selectin interaction and leukocyte function associated antigen (LFA)-  
 CC 1/intercellular adhesion molecule (ICAM) and very late antigen  
 CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
 CC includes an immunotherapeutic agent, which comprises a purified protein  
 CC extract that is isolated by diethylaminoethyl Sephadex chromatography of  
 CC a Nonidet P-40 insoluble particulate antigen fraction derived from  
 CC isolated killed cells of amastigotes from one or more species of the  
 CC Leishmania genus, where the particulate antigen fraction is solubilized  
 CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
 CC to diethylaminoethyl Sephadex and is eluted with a solution comprising  
 CC 0.1 M sodium chloride, 8 M urea and 0.025 M  
 CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract  
 CC includes polypeptides having apparent molecular weights of 73, 80 and 82  
 CC kDa, after total reduction and alkylation. The species is L. amazonensis,  
 CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the  
 CC invention has antipsoriatic and immunostimulant activity, and inhibits T-  
 CC cell rolling by interfering with CLA-E selectin interaction and LFA-  
 CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
 CC inhibiting T-cell rolling in a human host, and for treating psoriasis.  
 CC The present sequence represents an Leishmania peptide used in the  
 CC invention.  
 XX Sequence 9 AA;  
 SQ Query Match 77.8%; Score 35; DB 9; Length 9;  
 Best Local Similarity 77.8%; Pred. No. 2.1e+06;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AQVEDAIQK 9  
 DB 1 AQVEDAIQK 9  
 RESULT 5  
 AAR67587  
 ID AAR67587 standard; protein; 164 AA.  
 XX AAR67587;  
 AC AAR67587;  
 XX 06-OCT-1995 (first entry)  
 DT

XX Bovine male enhanced antigen (Mea).  
 DE Bovine male enhanced antigen; Mea; gender discrimination; cow; primers;  
 XX probes.  
 KW Bos taurus.  
 OS JP06319546-A.  
 XX 22-NOV-1994.  
 PN 07-MAY-1993; 93JP-00130055.  
 XX 07-MAY-1993; 93JP-00130055.  
 PR (KACH-) KACHIKU JUSEIRAN ISHOKU GIKUTSU KENKYUKU.  
 XX WPI; 1995-040314/06.  
 DR N-PSDB; AAQ79135, AAQ79136, AAQ79137.  
 XX DNA sequences from mice and cattle - used as primers and probes for the  
 PT discrimination of gender.  
 PS Claim 7; Page 9; 12pp; Japanese.  
 XX AAQ79136 encodes AAR67587 bovine male enhanced antigen (Mea), the cDNA  
 CC can be used to produce probes and primers for the discrimination of  
 CC genders from tissue samples and embryos  
 XX Sequence 164 AA;  
 SQ Query Match 77.8%; Score 35; DB 2; Length 164;  
 Best Local Similarity 66.7%; Pred. No. 92;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AQVEDAIQK 9  
 DB 144 AQWEDVVQK 152  
 RESULT 6  
 AD862272  
 ID ADE62272 standard; protein; 172 AA.  
 XX ADE62272;  
 AC ADE62272;  
 XX 29-JAN-2004 (first entry)  
 DT Human Protein Q16626, SEQ ID NO 8201.  
 XX Human; pain; neuronal tissue; gene therapy;  
 XX spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX Homo sapiens.  
 OS WO2003016475-A2.  
 PN 27-FEB-2003.  
 PD 14-AUG-2002; 2002WO-US025765.  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 DR

DR GENBANK; Q16626.

XX New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX

PS Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotides or a polynucleotide which represents a fragment,

CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell

CC comprising the vector, a method for identifying a nucleotide sequence

CC which is differentially regulated in an animal subjected to pain and a

CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence

CC that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates

CC the expression of a polynucleotide sequence which is differentially

CC expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the

CC polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more

CC polypeptides or their antibodies. The polynucleotide or the compound that

CC modulates its activity is useful for preparing a medicament for treating

CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene

CC therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic form directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 172 AA;

Query Match 77.8%; Score 35; DB 7; Length 172;

Best Local Similarity 66.7%; Pred. No. 97;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQYEDAIQK 9

Db 152 AQWEDVVQK 160

|||||

|||||

RESULT 7

AA067586

ID AAR67586 standard; protein; 174 AA.

AC AAR67586;

XX

XX 06-OCT-1995 (first entry)

DT

XX Murine male enhanced antigen (Mea).

DE

XX Murine male enhanced antigen; Mea; gender discrimination; mouse; primers;

KW probes.

XX

XX Mus musculus.

OS

XX JP06319546-A.

PN

XX 22-NOV-1994.

PD

XX 07-MAY-1993; 93JP-00130055.

PF

XX 07-MAY-1993; 93JP-00130055.

PR

XX (KACH-) KACHIKU JUSETRAN ISHOKU GIKUTSU KENKYUKU.

PA

XX WPI; 1995-040314/06.

DR N-PSDB; AAR67586.

XX DNA sequences from mice and cattle - used as primers and probes for the

PT discrimination of gender.

XX

PS Claim 2; Page 6-7; 12pp; Japanese.

XX

CC AA079134 encodes AAR67586 murine male enhanced antigen (Mea), the cDNA

CC can be used to produce probes and primers for the discrimination of

CC genders from tissue samples and embryos

XX

SQ Sequence 174 AA;

Query Match 77.8%; Score 35; DB 2; Length 174;

Best Local Similarity 66.7%; Pred. No. 98;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQYEDAIQK 9

Db 154 AQWEDVVQK 162

|||||

|||||

RESULT 8

ABU07424

ID ABU07424 standard; protein; 185 AA.

XX

AC ABU07424;

XX

XX 28-JAN-2003 (first entry)

DT

XX Protein differentially regulated in prostate cancer #27.

DE

XX Prostate cancer; gene expression; differential regulation;

KW molecular marker; drug target; cancer detection; cancer diagnosis;

KW cancer staging; cancer grading; cancer assessing; cancer monitoring.

XX

OS Homo sapiens.

XX

XX WO200281638-A2.

PN

XX 17-OCT-2002.

PD

XX 08-APR-2002; 2002WO-US010824.

PF

XX 06-APR-2001; 2001US-0281731P.

PR

XX 06-APR-2001; 2001US-0281732P.

XX

XX (ORIG-) ORIGENE TECHNOLOGIES INC.

PA

XX Sun Z, Jay G;

PI

XX WPI; 2003-058520/05.

DR

XX Novel genes which are differentially regulated in prostate cancer, useful

PT for diagnosing prostate cancer in prostate tissue sample and assessing

PT therapeutic or preventive intervention in prostate cancer patients.

XX

PS Claim 1; Page 262-263; 416pp; English.

XX

XX The invention describes genes (I) which are differentially regulated in

CC prostate cancer. (I) is useful for diagnosing a prostate cancer in a

CC sample comprising prostate tissue, which involves determining the number

CC of target genes which are differentially-regulated in the sample, where

CC the number is indicative of the probability that the sample comprises

CC prostate cancer. (I) is useful for assessing a therapeutic or preventive

CC intervention in a subject having a prostate cancer, which involves

CC determining the expression levels in a sample comprising prostate tissue

CC of target genes which are differentially-regulated in prostate cancer.

CC Preferably, the expression levels of at least 10 genes are determined.

CC (I) is also useful for identifying agents that modulate a biological

CC activity of a polypeptide differentially-regulated in prostate cancer

CC cells, which involves contacting a polypeptide differentially-regulated

CC in prostate cancer cells with a test agent under conditions effective for

CC the test agent to modulate a biological activity of the polypeptide, and



CC determining whether the test agent modulates the biological activity. (I)  
 CC is useful as molecular markers, as drug targets, and for detecting,  
 CC diagnosing, staging, grading, assessing, monitoring, prognosticating,  
 CC preventing or treating, determining predisposition to diseases and  
 CC conditions especially relating to prostate cancer. (I) and its expression  
 CC products are used in the diagnostic test to assay for presence of cancer  
 CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in  
 CC blood etc. (I) is useful for assessing cancer e.g., to determine the type  
 CC of cancer, its stage of development, the nature of genetic defect, etc.  
 CC The polypeptide encoded by (I) can be used as target for therapy or drug  
 CC discovery. (I) can also be used for expressing the polypeptide and thus  
 CC for searching specific binding partners of the polypeptide. (I) is useful  
 CC in therapeutic applications to treat prostate cancer. The identification  
 CC of specific genes, and groups of genes, expressed in pathways  
 CC physiologically relevant to prostate cancer permits the definition of  
 CC functional and disease pathways and the delineation of targets in these  
 CC pathways which are useful in diagnostic, therapeutic, and clinical  
 CC applications. This is the amino acid sequence of a protein differentially  
 CC regulated in prostate cancer

XX SQ Sequence 185 AA;

Query Match 77.8%; Score 35; DB 6; Length 185;  
 Best Local Similarity 66.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9  
 Db 165 AQWEDVVQK 173  
 ||:||:|

RESULT 9  
 ADX07373

ID ADX07373 standard; protein; 185 AA.

AC ADX07373;

XX 21-APR-2005 (first entry)

XX Cyclin-dependent kinase modulation biomarker SEQ ID NO 1938.

XX cytotstatic; cyclin-dependent kinase; cdk; biomarker.

XX Homo sapiens.

XX WO2005012875-A2.

XX 10-FEB-2005.

XX 29-JUL-2004; 2004WO-US024424.

XX 29-JUL-2003; 2003US-0490890P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;

XX WPI; 2005-163068/17.

XX N-PSDB; ADX07372.

XX Biomarkers useful for predicting or determining the response of a mammal  
 PT to a cancer treatment comprising administration of a modulator of cyclin-  
 PT dependent kinase activity.

PS Claim 5; SEQ ID NO 1938; 141pp; English.

XX This invention describes a novel method of predicting or determining  
 CC whether a mammal will respond or is responding to an anti-cancer agent  
 CC that modulates cyclin-dependent kinase (cdk) activity. The method  
 CC comprises measuring the level of one or more biomarkers selected from  
 CC 2774 biomarkers given in the specification (nucleotide sequence SEQ ID  
 CC NO:1246 (Genbank EST W28729) is especially preferred). The method of the  
 CC invention is utilized in a kit for determining or predicting whether

CC patient would be susceptible or resistant to treatment by an agent  
 CC modulating cdk activity. The invention also describes a method for  
 CC utilizing individualized genetic profiles for treating diseases and  
 CC disorders based on patient's response and molecular level, specialized  
 CC microarrays comprising the biomarkers described, antibodies directed  
 CC against the biomarkers and a cell culture model to identify biomarkers.  
 CC The cdk modulator is preferably N-5-([5-(1,1-Dimethylethyl)-2-oxazolyl]  
 CC oxazolyl][methylthio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-  
 CC tartaric acid salt. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at fip.wipo.int/pub/published pct sequences. This  
 CC sequence represents a biomarker used in the method of the invention.

XX SQ Sequence 185 AA;

Query Match 77.8%; Score 35; DB 9; Length 185;  
 Best Local Similarity 66.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9  
 Db 165 AQWEDVVQK 173  
 ||:||:|

RESULT 10

AEAI5501  
 ID AEA15501 standard; protein; 644 AA.

XX AEA15501;

XX 28-JUL-2005 (first entry)

XX Human polypeptide #120.

XX Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
 KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
 KW cytostatic; neoplasm.

XX Homo sapiens.

XX WO2005047534-A2.

XX 26-MAY-2005.

XX 15-OCT-2004; 2004WO-EP011599.

XX 28-OCT-2003; 2003EP-00024565.

XX (FARB ) BAYER HEALTHCARE AG.

XX Wirtz R, Munnes M;

XX WPI; 2005-372393/38.

XX N-PSDB; AEA15423.

XX Predicting a response to cancer treatment by detecting at least 2  
 PT markers, which are genes or genomic nucleic acid sequences that are  
 PT located on one chromosomal region, which is altered in malignant  
 PT neoplasia.

XX Claim 7; SEQ ID NO 463; 464pp; English.

XX The invention relates to a method of predicting response to cancer  
 CC treatment comprising detection of at least 2 markers, where the markers  
 CC are genes and fragments or genomic nucleic acid sequences that are  
 CC located on one chromosomal region, which is altered in malignant  
 CC neoplasia. The invention also relates to a method for the prediction,  
 CC diagnosis or prognosis of malignant neoplasia, methods for detecting,  
 CC deregulations in malignant neoplasia and breast cancer, a method of  
 CC determining the phenotype of a cell or tissue, a method for identifying  
 CC genomic regions which are altered on the chromosomal level and encode  
 CC genes that are linked by function and are differentially expressed in  
 CC malignant neoplasia and breast cancer, methods of screening for agents

CC which regulate the activity of a polypeptide or a polynucleotide and  
CC antibodies that specifically bind to a full length or partial  
CC polypeptide. The method is useful for predicting response to cancer  
CC treatment. The methods and compositions are useful for predicting,  
CC diagnosing, prognosing, preventing or treating malignant neoplasia  
CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
CC lung cancer. This sequence represents a human polypeptide used in the  
CC scope of the invention.

XX  
SQ Sequence 644 AA;

Query Match 77.8%; Score 35; DB 9; Length 644;  
Best Local Similarity 77.8%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
|||||

DB 356 AQYEDAIQK 364

## RESULT 11

ABB48335  
ID ABB48335 standard; protein; 785 AA.

XX  
AC ABB48335;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #1039.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR001118.

XX 11-APR-2000; 2000FR-00004629.

XX (INSP ) INST PASTEUR.

XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;  
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tietz-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;

XX WPI; 2002-010914/01.

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and related  
PT polypeptides.

XX Claim 6; SEQ ID NO 1040; 192pp; French.

XX The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present invention is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies. Identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for

CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccine compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 785 AA;

Query Match 77.8%; Score 35; DB 5; Length 785;  
Best Local Similarity 75.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9

DB 716 QYEDAIQK 723

## RESULT 12

ABU32679

ID ABU32679 standard; protein; 785 AA.

XX  
AC ABU32679;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #18206.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Listeria monocytogenes.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.  
XX N-PSDB; ACA36549.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 60603; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 785 AA;

Query Match 77.8%; Score 35; DB 6; Length 785;

Best Local Similarity 75.0%; Pred. No. 5.3e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9

Db 716 RYEDALQK 723

RESULT 13

ID ADN19018 standard; protein; 875 AA.

XX AC ADN19018;

DT 02-DEC-2004 (first entry)

DE Bacterial polypeptide #1671.

XX Recombinant DNA construct; transformed plant; improved plant property;  
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
 KW pathogen tolerance; pest tolerance; plant disease resistance;  
 KW cell cycle pathway modification; plant growth regulator;  
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 KW bacterial polypeptide.

XX Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 1671; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polypeptide used in the  
 CC scope of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 875 AA;

Query Match 77.8%; Score 35; DB 8; Length 875;

Best Local Similarity 66.7%; Pred. No. 6e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9

Db 529 AQYEDLVKK 537

RESULT 14

ID AAM99474

XX AAM99474 standard; peptide; 9 AA.

XX AC AAM99474;

XX 07-DEC-2001 (first entry)

XX Vaccine related MHC ligand peptide SEQ ID NO:577.

XX Glutamic acid; glutamine; vaccine; major histocompatibility complex; MHC;  
 KW immunomodulator; anti-allergic; endocrine; neuroprotectant; virucidal;  
 KW bactericidal; antiparasitic; fungicidal; cytostatic; medicine;  
 KW pharmaceutical; immune disorder; immune deficiency; autoimmune;  
 KW hypersensitivity; allergy; graft rejection; infection; hormonal disorder;  
 KW central nervous system disease; cancer; melanoma; anti-melanoma vaccine;  
 KW human immunodeficiency virus.

XX Homo sapiens.

XX WO200170772-A2.

XX 27-SEP-2001.

XX 22-MAR-2001; 2001WO-FR000872.

XX 23-MAR-2000; 2000FR-00003711.

XX (FABR-) FABRE MEDICAMENT SA PIERRE.

XX Klinguer-Hamur C, Corvaia N, Beck A, Goetsch L;

XX WPI; 2001-611470/70.

XX Stabilized pharmaceutical containing N-terminal glutamic acid or  
 PT glutamine, useful e.g. in anti-melanoma vaccines, is an addition salt

PT with strong acid.  
 XX Claim 9; Page 129; 149pp; French.  
 PS  
 XX The present invention describes a pharmaceutical compound (I) that  
 CC contains an N-terminal glutamic acid (Glu) or glutamine (Gln) residue in  
 CC the form of an addition salt with a strong, physiologically acceptable  
 CC acid (II). Also described are: (a) a pharmaceutical composition  
 CC containing at least one (I); (b) a vaccine containing at least one (I)  
 CC where this is a major histocompatibility complex (MHC) ligand (Ia); (c) a  
 CC method for in vitro diagnosis of diseases associated with the presence of  
 CC (Ia); (d) a kit for method (c) that includes a (Ia); and (e) a process  
 CC for preparing (I). (I) has immunomodulator, endocrine, antiallergic,  
 CC neuroprotectant, virucidal, bactericidal, antiparasitic, fungicidal and  
 CC cytostatic activities. (I) are useful, in human or veterinary medicine,  
 CC in pharmaceutical compositions (for treating immune disorders, e.g.  
 CC immune deficiency, autoimmune states, hypersensitivity, allergy, graft  
 CC rejection, infection, hormonal disorders and central nervous system  
 CC diseases), also, where (I) is a MHC ligand (Ia), in vaccines for  
 CC treatment or prevention of: (i) viral, bacterial, parasitic or fungal  
 CC infections; or (ii) of cancers. A particular application is in anti-  
 CC melanoma vaccines. (I) are also useful for in vitro diagnosis of diseases  
 CC associated with interactions between MHC and (I), e.g. melanoma and human  
 CC immunodeficiency virus infection. AAM98898 to AAM99592 represent peptides  
 CC which can be used in pharmaceutical compounds from the present invention  
 XX  
 XX Sequence 9 AA;  
 SQ  
 Query Match 75.6%; Score 34; DB 4; Length 9;  
 Best Local Similarity 62.5%; Pred. No. 2.1e+06;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QYEDAIQK 9  
 Db ||:|:|:|  
 1 QYDDAVEK 8  
 RESULT 15  
 ADH12886  
 ID ADH12886 standard; protein; 124 AA.  
 XX  
 AC ADH12886;  
 XX  
 DT 25-MAR-2004 (first entry)  
 XX  
 DE Francisella tularensis immunogenic protein 67, SEQ ID NO:67.  
 XX  
 KW Immunogenic protein; protective immune response; vaccine;  
 KW genetic vaccine; antibacterial.  
 XX  
 OS Francisella tularensis.  
 XX  
 PN WO2004003009-A2.  
 XX  
 PD 08-JAN-2004.  
 XX  
 PF 26-JUN-2003; 2003WO-CB002718.  
 XX  
 PR 28-JUN-2002; 2002GB-00014942.  
 XX  
 PA (MINA ) UK SEC FOR DEFENCE.  
 XX  
 PI Titball RW, Mayers CN, Duffield ML, Miller J, Rowe SC;  
 XX  
 DR WPI: 2004-083016/08.  
 DR N-PSDB; ADH12886.  
 XX  
 XX New protein, useful as a vaccine for producing a protective immune  
 PT response in a mammal against infection by Francisella tularensis, or  
 PT preventing or treating Francisella tularensis infection in a mammal.  
 XX  
 PS Claim 1; SEQ ID NO 67; 217pp; English.  
 XX  
 CC The invention relates to 100 potentially immunogenic proteins from  
 CC Francisella tularensis (ADH12820-ADH12919) and nucleic acids encoding  
 CC them (ADH12920-ADH13019) which may be used in vaccines. The nucleic acids  
 CC may be contained within a vaccine, bacterial or plasmid vector. The  
 CC invention also relates to a pharmaceutical composition comprising the  
 CC protein or the nucleic acid in combination with a pharmaceutical carrier  
 CC or excipient. The proteins and nucleic acids are useful in vaccine  
 CC compositions for producing a protective immune response against  
 CC Francisella tularensis infection in a mammal, or for preventing or  
 CC treating Francisella tularensis infection in a mammal. The present  
 CC sequence represents a Francisella tularensis protein of the invention.  
 XX  
 XX Sequence 124 AA;  
 SQ  
 Query Match 75.6%; Score 34; DB 8; Length 124;  
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AQYEDAIQ 8  
 Db |||||  
 76 AQYESAIQ 83  
 Search completed: July 5, 2006, 18:43:47  
 Job time : 80.2039 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 65.8835 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-4

Perfect score: 45

Sequence: 1 AQYEDAIQK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	100.0	9	5	US-10-774-928-4
2	35	77.8	9	5	US-10-774-928-2
3	35	77.8	258	5	US-10-732-923-14076
4	35	77.8	785	4	US-10-282-122A-60603
5	35	77.8	785	6	US-11-045-004-1040
6	35	77.8	875	4	US-10-369-493-1671
7	34	75.6	9	4	US-10-239-313A-577
8	34	75.6	77	4	US-10-424-599-214479
9	34	75.6	148	4	US-10-767-701-57385
10	34	75.6	233	3	US-09-815-242-10642
11	34	75.6	233	4	US-10-282-122A-57284
12	34	75.6	366	4	US-10-437-963-181853
13	33	73.3	117	4	US-10-108-260A-2985
14	33	73.3	396	4	US-10-179-766-7
15	33	73.3	530	4	US-10-437-963-197053
16	33	73.3	590	4	US-10-275-595A-33
17	33	73.3	591	4	US-10-437-963-197054
18	33	73.3	808	4	US-10-282-122A-52904
19	33	73.3	908	4	US-10-282-122A-47185
20	33	73.3	1132	4	US-10-437-963-160810
21	32	71.1	10	6	US-11-129-104-74
22	32	71.1	60	4	US-10-425-115-186946
23	32	71.1	129	5	US-10-737-288-13
24	32	71.1	129	5	US-10-891-825-13
25	32	71.1	148	5	US-10-491-467-25
26	32	71.1	149	5	US-10-491-467-26
27	32	71.1	162	4	US-10-767-701-52156

28 71.1 173 3 US-09-918-715-248 Sequence 248, Appli  
29 71.1 173 4 US-10-427-887-3 Sequence 3, Appli  
30 71.1 173 4 US-10-427-887-6 Sequence 6, Appli  
31 71.1 173 4 US-10-444-795B-799 Sequence 799, App  
32 71.1 173 4 US-10-474-794-248 Sequence 248, App  
33 71.1 173 5 US-10-979-159-248 Sequence 248, App  
34 71.1 173 6 US-11-129-104-88 Sequence 88, Appli  
35 71.1 199 4 US-10-767-701-35473 Sequence 35473, A  
36 71.1 200 4 US-10-437-963-140077 Sequence 140077,  
37 71.1 221 4 US-10-291-190-36 Sequence 36, Appli  
38 71.1 229 4 US-10-425-115-193806 Sequence 193806,  
39 71.1 235 6 US-11-129-104-90 Sequence 90, Appli  
40 71.1 239 4 US-10-335-977-8903 Sequence 8903, Ap  
41 71.1 243 6 US-11-129-104-91 Sequence 91, Appli  
42 71.1 293 4 US-10-163-214-10 Sequence 10, Appli  
43 71.1 358 4 US-10-727-936-38 Sequence 38, Appli  
44 71.1 385 4 US-10-369-493-3196 Sequence 3196, Ap  
45 71.1 405 4 US-10-335-977-8905 Sequence 8905, Ap

## ALIGNMENTS

## RESULT 1

US-10-774-928-4  
; Sequence 4, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-4

Query Match 100.0%; Score 45; DB 5; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.9e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
DB 1 AQYEDAIQK 9

## RESULT 2

US-10-774-928-2  
; Sequence 2, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-2

Query Match 77.8%; Score 35; DB 5; Length 9;  
Best Local Similarity 77.8%; Pred. No. 1.9e+06;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY      1 AQYEDAIQK 9
      |||||
DB      1 AQYEDAIQK 9

RESULT 3
US-10-732-923-14076
; Sequence 14076, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 14076
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Encephalitozoon cuniculi
US-10-732-923-14076

Query Match      77.8%; Score 35; DB 5; Length 258;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 QYEDAIQK 9
      |||||
DB      5 QYEEALQK 12

RESULT 4
US-10-282-122A-60603
; Sequence 60603, Application US/10282122A
; Publication No. US2004029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60603
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-10-282-122A-60603

Query Match      77.8%; Score 35; DB 4; Length 785;
Best Local Similarity 75.0%; Pred. No. 5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 QYEDAIQK 9
      |||||
DB      716 RYEDALQK 723

RESULT 5
US-11-045-004-1040
; Sequence 1040, Application US/11045004
; Publication No. US20060078901A1
; GENERAL INFORMATION:
; APPLICANT: BUCHRIESER, CARMEN
; APPLICANT: FRANGEUL, LIONEL
; APPLICANT: COUVE, ELISABETH
; APPLICANT: RUSNIOK, CHRISTOPHE
; APPLICANT: FSIHI, HAFIDA
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DUSSURGET, OLIVIER
; APPLICANT: CHETOUANI, FARID
; APPLICANT: NEDJARI, HAFED
; APPLICANT: GLASER, PHILIPPE
; APPLICANT: KUNST, FRANCK
; APPLICANT: COSSART, PASCALE
; APPLICANT: DANIELS, JUSTIN
; APPLICANT: GORBEL, WERNER
; APPLICANT: KREFT, JURGEN
; APPLICANT: KUHN, MICHAEL
; APPLICANT: NG, EVA
; APPLICANT: VAZQUEZ-BOLAND, ANTONIO
; APPLICANT: DOMINGUEZ-BERNAL, GUSTAVO
; APPLICANT: GARRIDO-GARCIA, PATRICIA
; APPLICANT: TIERREZ-MARTINEZ, ALBERTO
; APPLICANT: AMEND, ALEXANDRA
; APPLICANT: CHAKRABORTY, TRINAD
; APPLICANT: DOWANN, EUGEN
; APPLICANT: HAIN, THORSTEN
; APPLICANT: BERCHE, PATRICK
; APPLICANT: CHARBIT, ALAIN
; APPLICANT: DURANT, LIONEL
; APPLICANT: PEREZ-DIAZ, JOSE-CLAUDIO
; APPLICANT: BAQUERO, FERNANDO
; APPLICANT: GARCIA DEL PORTILLO, FRANCISCO
; APPLICANT: GOMEZ-LOPEZ, NURIA
; APPLICANT: MADUENIO, ENCARN
; APPLICANT: PABLOS, BETRIZ DE
; APPLICANT: WEHLAND, JURGEN
; APPLICANT: KARST, UWE
; APPLICANT: ENTIAN, KARL-DIETER
; APPLICANT: HAUF, JORG
; APPLICANT: ROSE, MATTHIAS
; APPLICANT: VOSS, HAMUT
; TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
; FILE REFERENCE: 05394.0018-02
; CURRENT APPLICATION NUMBER: US/11/045,004
; CURRENT FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: 10/637,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: 10/257,023
; PRIOR FILING DATE: 2002-10-08
```

; PRIOR APPLICATION NUMBER: PCT/FR01/01118  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: FR 00/04,629  
; PRIOR FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 2854  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 1040  
; LENGTH: 785  
; TYPE: PRT  
; ORGANISM: Listeria monocytogenes  
US-11-045-004-1040

Query Match 77.8%; Score 35; DB 6; Length 785;  
Best Local Similarity 75.0%; Pred. No. 5e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9  
Db 716 RYEDALQK 723  
:||||:|

## RESULT 6

US-10-369-493-1671  
; Sequence 1671, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 1671  
; LENGTH: 875  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1671

Query Match 77.8%; Score 35; DB 4; Length 875;  
Best Local Similarity 66.7%; Pred. No. 5.6e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
Db 529 AQYEDLVKK 537  
:||||:|

## RESULT 7

US-10-239-313A-577  
; Sequence 577, Application US/10239313A  
; Publication No. US20030175285A1  
; GENERAL INFORMATION:  
; APPLICANT: KLINGUER - HAMOUR, Christine  
; APPLICANT: CORVAIA, Nathalie  
; APPLICANT: BECK, Alain  
; APPLICANT: GOETSCH, Liliane  
; TITLE OF INVENTION: MOLECULE OF PHARMACEUTICAL INTEREST COMPRISING AT ITS  
; FILE REFERENCE: N-TERMINAL A GLUTAMIC ACID OR A GLUTAMINE IN THE FORM  
; TITLE OF INVENTION: OF A PHYSIOLOGICALLY ACCEPTABLE STRONG ACID  
; FILE REFERENCE: 343 727 - US  
; CURRENT APPLICATION NUMBER: US/10/239,313A  
; CURRENT FILING DATE: 2002-09-19  
; PRIOR APPLICATION NUMBER: FR 00/03711  
; PRIOR FILING DATE: 2000-03-23  
; PRIOR APPLICATION NUMBER: PCT 01/70772  
; PRIOR FILING DATE: 2001-03-22

; NUMBER OF SEQ ID NOS: 697  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 577  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-239-313A-577

Query Match 75.6%; Score 34; DB 4; Length 9;  
Best Local Similarity 62.5%; Pred. No. 1.9e+06;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9  
Db 1 QYDDAVEK 8  
:||||:|

## RESULT 8

US-10-424-599-214479  
; Sequence 214479, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 214479  
; LENGTH: 77  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(77)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_35701C.1.pap  
US-10-424-599-214479

Query Match 75.6%; Score 34; DB 4; Length 77;  
Best Local Similarity 66.7%; Pred. No. 65;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
Db 67 AQHEDALQK 75  
:||||:|

## RESULT 9

US-10-767-701-57385  
; Sequence 57385, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 57385  
; LENGTH: 148  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 30966480.pap

US-10-767-701-57385

Query Match 75.6%; Score 34; DB 4; Length 148;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YEDAIQK 9  
|||||  
Db 67 YEDALQK 73

RESULT 10

US-09-815-242-10642  
; Sequence 10642, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10642  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis

US-09-815-242-10642

Query Match 75.6%; Score 34; DB 3; Length 233;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YEDAIQK 9  
|||||  
Db 36 YEDALQK 42

RESULT 11

US-10-282-122A-57284  
; Sequence 57284, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Lianguo  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 57284  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis

US-10-282-122A-57284

Query Match 75.6%; Score 34; DB 4; Length 233;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YEDAIQK 9  
|||||  
Db 36 YEDALQK 42

RESULT 12

US-10-437-963-181853  
; Sequence 181853, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 181853  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_79094C.1.pep

US-10-437-963-181853



Query Match 75.6%; Score 34; DB 4; Length 366;  
Best Local Similarity 77.8%; Pred. No. 3.4e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9  
DB 244 AQKEDAIQK 252  
|||:|

## RESULT 13

US-10-108-260A-2985  
; Sequence 2985, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2985  
; LENGTH: 117  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-2985

Query Match 73.3%; Score 33; DB 4; Length 117;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9  
DB 49 AQVEDAIQK 57  
|||:|

## RESULT 14

US-10-179-766-7  
; Sequence 7, Application US/10179766  
; Publication No. US20030190312A1  
; GENERAL INFORMATION:  
; APPLICANT: Kenyon, Cynthia  
; APPLICANT: Apfeld, Javier  
; APPLICANT: Dillin, Andrew  
; APPLICANT: Garigan, Delia  
; APPLICANT: Hsu, Ao-Lin A.  
; APPLICANT: Lehrer-Graiwer, Josh  
; APPLICANT: Murphy, Coleen  
; TITLE OF INVENTION: The Regents of the University of California  
; FILE REFERENCE: 023070-119930US  
; CURRENT APPLICATION NUMBER: US/10/179,766  
; CURRENT FILING DATE: 2002-06-24  
; PRIOR APPLICATION NUMBER: US 60/300,577  
; PRIOR FILING DATE: 2001-06-22  
; PRIOR APPLICATION NUMBER: US 60/301,052  
; PRIOR FILING DATE: 2001-06-25  
; PRIOR APPLICATION NUMBER: US 60/373,975  
; PRIOR FILING DATE: 2002-04-18  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 396  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
; OTHER INFORMATION: llw-1 (long-lived worm protein), gene Y54G11A.8  
US-10-179-766-7

Query Match 73.3%; Score 33; DB 4; Length 396;  
Best Local Similarity 85.7%; Pred. No. 5.8e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 8  
DB 108 QYDDAIQK 114  
|||:|

## RESULT 15

US-10-437-963-197053  
; Sequence 197053, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 197053  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_92848C.1.pap  
US-10-437-963-197053

Query Match 73.3%; Score 33; DB 4; Length 530;  
Best Local Similarity 66.7%; Pred. No. 7.9e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9  
DB 491 AAYEDTLQK 499  
|||:|

Search completed: July 5, 2006, 20:51:57  
Job time : 66.8835 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 5.15534 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-4

Perfect score: 45

Sequence: 1 AQVEDAIQK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:\*

1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB pep.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	75.6	447	6	US-10-449-902-51639
2	33	73.3	117	7	US-11-293-697-2985
3	33	73.3	591	6	US-10-449-902-53151
4	32	71.1	173	6	US-10-537-303-10
5	32	71.1	321	6	US-10-449-902-41051
6	31	68.9	167	6	US-10-537-303-4
7	31	68.9	167	6	US-10-537-303-6
8	31	68.9	173	6	US-10-537-303-2
9	31	68.9	250	6	US-10-449-902-32999
10	31	68.9	351	6	US-10-449-902-46480
11	31	68.9	691	6	US-10-953-349-1322
12	31	68.9	705	6	US-10-953-349-1321
13	31	68.9	713	6	US-10-953-349-1320
14	31	68.9	723	6	US-10-449-902-53379
15	31	68.9	809	6	US-10-449-902-54024
16	30	66.7	23	7	US-11-134-871-3360
17	30	66.7	103	6	US-10-953-349-26513
18	30	66.7	324	6	US-10-953-349-25046
19	30	66.7	329	6	US-10-953-349-25045
20	30	66.7	330	6	US-10-953-349-25044
21	30	66.7	335	6	US-10-953-349-32983
22	30	66.7	351	6	US-10-953-349-32982
23	30	66.7	462	6	US-10-196-749-298
24	30	66.7	553	6	US-10-449-902-52615
25	30	66.7	1180	6	US-10-505-928-459

26	64.4	29	64.4	130	6	US-10-953-349-525	Sequence 525, App
27	64.4	29	64.4	161	6	US-10-953-349-19502	Sequence 19502, A
28	64.4	29	64.4	173	6	US-10-953-349-19501	Sequence 19501, A
29	64.4	29	64.4	181	7	US-11-174-307B-5276	Sequence 5276, Ap
30	64.4	29	64.4	232	6	US-10-953-349-19523	Sequence 19523, A
31	64.4	29	64.4	265	6	US-10-953-349-27010	Sequence 27010, A
32	64.4	29	64.4	279	6	US-10-449-902-49131	Sequence 49131, A
33	64.4	29	64.4	297	6	US-10-953-349-19522	Sequence 19522, A
34	64.4	29	64.4	317	6	US-10-953-349-27009	Sequence 27009, A
35	64.4	29	64.4	347	6	US-10-953-349-19521	Sequence 19521, A
36	64.4	29	64.4	349	6	US-10-953-349-20252	Sequence 20252, A
37	64.4	29	64.4	467	6	US-10-953-349-20251	Sequence 20251, A
38	64.4	29	64.4	485	6	US-10-953-349-25060	Sequence 25060, A
39	64.4	29	64.4	490	6	US-10-953-349-20250	Sequence 20250, A
40	64.4	29	64.4	500	6	US-10-953-349-25059	Sequence 25059, A
41	64.4	29	64.4	507	6	US-10-953-349-25058	Sequence 25058, A
42	64.4	29	64.4	599	6	US-10-449-902-53037	Sequence 53037, A
43	64.4	29	64.4	627	6	US-10-471-571A-3090	Sequence 3090, Ap
44	64.4	29	64.4	649	6	US-10-449-902-37545	Sequence 37545, A
45	64.4	29	64.4	732	6	US-10-756-047-2	Sequence 2, Appl

## ALIGNMENTS

## RESULT 1

US-10-449-902-51639  
; Sequence 51639, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205V1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51639  
; LENGTH: 447  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-51639

Query Match 75.6%; Score 34; DB 6; Length 447;  
Best Local Similarity 77.8%; Pred. No. 19;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQVEDAIQK 9

Db 364 AQEDALQK 372

## RESULT 2

US-11-293-697-2985  
; Sequence 2985, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 2985
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-2985

Query Match      73.3%; Score 33; DB 7; Length 117;
Best Local Similarity 77.8%; Pred. No. 1;
Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 0;

QY 1 AQYEDAIQK 9
Db 49 AQYERAIOE 57

RESULT 3
US-10-449-902-53151
; Sequence 53151, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A02051-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53151
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53151

Query Match      73.3%; Score 33; DB 6; Length 591;
Best Local Similarity 66.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9
Db 552 AAYEDTLQK 560

RESULT 4
US-10-537-303-10
; Sequence 10, Application US/10537303
; Publication No. US20060135419A1
; GENERAL INFORMATION:
; APPLICANT: Meise, Martin
; APPLICANT: Eulenber, Karsten
; APPLICANT: Nguyen, Tri
; APPLICANT: Tssetsenis, Theodoros
; TITLE OF INVENTION: Proteins Involved in the Regulation of Energy Homeostasis
; FILE REFERENCE: 2923-703
; CURRENT APPLICATION NUMBER: US/10/537,303
; CURRENT FILING DATE: 2005-06-02
; PRIOR APPLICATION NUMBER: PCT/EP03/013655
; PRIOR FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: EP 02 026 921.3
; PRIOR FILING DATE: 2002-12-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-537-303-10

Query Match      71.1%; Score 32; DB 6; Length 173;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8
Db 125 KYEDAIQ 131

RESULT 5
US-10-449-902-41051
; Sequence 41051, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A02051-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41051
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-41051

Query Match      71.1%; Score 32; DB 6; Length 321;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQK 9
Db 181 KYEEAVQK 188

RESULT 6
US-10-537-303-4
; Sequence 4, Application US/10537303
; Publication No. US20060135419A1
; GENERAL INFORMATION:
; APPLICANT: Meise, Martin
; APPLICANT: Eulenber, Karsten
; APPLICANT: Nguyen, Tri
; APPLICANT: Tssetsenis, Theodoros
; TITLE OF INVENTION: Proteins Involved in the Regulation of Energy Homeostasis
; FILE REFERENCE: 2923-703
; CURRENT APPLICATION NUMBER: US/10/537,303
; CURRENT FILING DATE: 2005-06-02
; PRIOR APPLICATION NUMBER: PCT/EP03/013655
; PRIOR FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: EP 02 026 921.3
; PRIOR FILING DATE: 2002-12-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-537-303-4

Query Match      68.9%; Score 31; DB 6; Length 167;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

QY 2 QYEDAIQ 8  
Db 122 KYEDAVQ 128

RESULT 7  
US-10-537-303-6  
; Sequence 6, Application US/10537303  
; Publication No. US20060135419A1  
; GENERAL INFORMATION:  
; APPLICANT: Meise, Martin  
; APPLICANT: Eulenber, Karsten  
; APPLICANT: Nguyen, Tri  
; APPLICANT: Tsetsenis, Theodoros  
; TITLE OF INVENTION: Proteins Involved in the Regulation of Energy Homeostasis  
; FILE REFERENCE: 2923-703  
; CURRENT APPLICATION NUMBER: US/10/537,303  
; CURRENT FILING DATE: 2005-06-02  
; PRIOR APPLICATION NUMBER: PCT/EP03/013655  
; PRIOR FILING DATE: 2003-12-03  
; PRIOR APPLICATION NUMBER: EP 02 026 921.3  
; PRIOR FILING DATE: 2002-12-03  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6  
; LENGTH: 167  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-537-303-6

Query Match 68.9%; Score 31; DB 6; Length 167;  
Best Local Similarity 71.4%; Pred. No. 26;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8  
Db 122 KYEDAVQ 128

RESULT 8  
US-10-537-303-2  
; Sequence 2, Application US/10537303  
; Publication No. US20060135419A1  
; GENERAL INFORMATION:  
; APPLICANT: Meise, Martin  
; APPLICANT: Eulenber, Karsten  
; APPLICANT: Nguyen, Tri  
; APPLICANT: Tsetsenis, Theodoros  
; TITLE OF INVENTION: Proteins Involved in the Regulation of Energy Homeostasis  
; FILE REFERENCE: 2923-703  
; CURRENT APPLICATION NUMBER: US/10/537,303  
; CURRENT FILING DATE: 2005-06-02  
; PRIOR APPLICATION NUMBER: PCT/EP03/013655  
; PRIOR FILING DATE: 2003-12-03  
; PRIOR APPLICATION NUMBER: EP 02 026 921.3  
; PRIOR FILING DATE: 2002-12-03  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2  
; LENGTH: 173  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-537-303-2

Query Match 68.9%; Score 31; DB 6; Length 173;  
Best Local Similarity 71.4%; Pred. No. 27;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8  
Db 125 KYEDAVQ 131

RESULT 9  
US-10-449-902-32999  
; Sequence 32999, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32999  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-32999

Query Match 68.9%; Score 31; DB 6; Length 250;  
Best Local Similarity 44.4%; Pred. No. 41;  
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 AOYEDAIQK 9  
Db 10 ABYQEAVER 18

RESULT 10  
US-10-449-902-46480  
; Sequence 46480, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46480  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-46480

Query Match 68.9%; Score 31; DB 6; Length 351;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AOYEDAI 7  
Db 251 AXYEDAI 257

RESULT 11  
US-10-953-349-1322  
; Sequence 1322, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:

```
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1322
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1322

Query Match      68.9%; Score 31; DB 6; Length 691;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QYEDAIOK 9
      |||| :||
Db      315 QYEDEVK 322

RESULT 12
US-10-953-349-1321
; Sequence 1321, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1321
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1321

Query Match      68.9%; Score 31; DB 6; Length 705;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QYEDAIOK 9
      |||| :||
Db      329 QYEDEVK 336

RESULT 13
US-10-953-349-1320
; Sequence 1320, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1320
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1320

Query Match      68.9%; Score 31; DB 6; Length 713;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      2 QYEDAIOK 9
      |||| :||
Db      337 QYEDEVK 344

RESULT 14
US-10-449-902-53379
; Sequence 53379, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53379
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53379

Query Match      68.9%; Score 31; DB 6; Length 723;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QYEDAIOK 9
      |||| :||
Db      344 QYEDEVK 351

RESULT 15
US-10-449-902-54024
; Sequence 54024, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54024
; LENGTH: 809
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-54024

Query Match      68.9%; Score 31; DB 6; Length 809;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QYEDAIOK 9
      |||| :||
Db      781 QYQAVTK 788
```

Search completed: July 5, 2006, 20:53:01  
Job time : 6.15534 secs

20060705

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 20.2718 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-4  
Perfect score: 45  
Sequence: 1 AQYEDAIQK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/iaa/5 COMB.pcp:\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/iaa/6 COMB.pcp:\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/iaa/7 COMB.pcp:\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/iaa/H COMB.pcp:\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/iaa/PCTUS COMB.pcp:\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/iaa/RE COMB.pcp:\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/iaa/backfiles.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	9	2 US-09-809-003A-4	Sequence 4, Appli
2	35	77.8	9	2 US-09-809-003A-2	Sequence 2, Appli
3	35	77.8	637	2 US-09-949-016-8152	Sequence 8152, Ap
4	33	73.3	390	2 US-09-248-796A-17724	Sequence 17724, A
5	33	73.3	527	2 US-09-252-991A-16954	Sequence 16954, A
6	32	71.1	173	2 US-08-725-532A-3	Sequence 3, Appli
7	32	71.1	173	2 US-09-164-193-2	Sequence 2, Appli
8	32	71.1	173	2 US-09-221-448A-2	Sequence 2, Appli
9	32	71.1	293	2 US-10-163-214-10	Sequence 10, Appl
10	32	71.1	315	2 US-09-543-681A-6875	Sequence 6875, Ap
11	32	71.1	358	2 US-09-889-463A-38	Sequence 38, Appl
12	32	71.1	424	2 US-09-328-352-6614	Sequence 6614, Ap
13	32	71.1	812	2 US-09-689-065B-6	Sequence 6, Appli
14	32	71.1	812	2 US-10-210-296A-6	Sequence 6, Appli
15	32	71.1	914	2 US-10-163-214-12	Sequence 12, Appl
16	32	71.1	1729	2 US-09-134-000C-5675	Sequence 5675, Ap
17	31	68.9	29	2 US-09-270-767-56905	Sequence 56905, A
18	31	68.9	103	2 US-09-198-452A-1230	Sequence 1230, Ap
19	31	68.9	130	2 US-09-270-767-59805	Sequence 59805, A
20	31	68.9	167	2 US-08-725-532A-5	Sequence 5, Appli
21	31	68.9	170	2 US-08-725-532A-1	Sequence 1, Appli
22	31	68.9	173	2 US-08-725-532A-6	Sequence 6, Appli
23	31	68.9	173	2 US-09-164-193-20	Sequence 20, Appl
24	31	68.9	173	2 US-09-221-448A-20	Sequence 20, Appl
25	31	68.9	173	2 US-09-949-016-6368	Sequence 6368, Ap
26	31	68.9	175	2 US-09-949-016-8724	Sequence 8724, Ap

27	31	68.9	176	2 US-09-270-767-31748	Sequence 31748, A
28	31	68.9	176	2 US-09-270-767-46965	Sequence 46965, A
29	31	68.9	177	2 US-09-438-185A-867	Sequence 867, App
30	31	68.9	194	2 US-09-949-016-10907	Sequence 10907, A
31	31	68.9	227	2 US-09-252-991A-31164	Sequence 31164, A
32	31	68.9	230	2 US-09-270-767-41662	Sequence 41662, A
33	31	68.9	275	2 US-09-468-433C-24	Sequence 24, Appl
34	31	68.9	312	2 US-09-270-767-33539	Sequence 33539, A
35	31	68.9	312	2 US-09-270-767-48756	Sequence 48756, A
36	31	68.9	330	2 US-09-270-767-61311	Sequence 61311, A
37	31	68.9	451	2 US-09-270-767-45789	Sequence 45789, A
38	31	68.9	544	2 US-09-328-352-4446	Sequence 4446, Ap
39	31	68.9	581	2 US-10-104-047-3540	Sequence 3540, Ap
40	31	68.9	663	2 US-09-248-796A-19801	Sequence 19801, A
41	30	66.7	85	2 US-09-543-681A-5028	Sequence 5028, Ap
42	30	66.7	86	2 US-09-107-532A-5983	Sequence 5983, Ap
43	30	66.7	324	2 US-09-902-540-15215	Sequence 15215, A
44	30	66.7	355	2 US-09-328-352-5122	Sequence 5122, Ap
45	30	66.7	363	2 US-09-270-767-33628	Sequence 33628, A

ALIGNMENTS

RESULT 1  
US-09-809-003A-4  
; Sequence 4, Application US/098090003A  
; Patent No. 6673351  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: Psoriasis  
; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-09-809-003A-4

Query Match 100.0%; Score 45; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIQK 9  
Db 1 AQYEDAIQK 9

RESULT 2  
US-09-809-003A-2  
; Sequence 2, Application US/098090003A  
; Patent No. 6673351  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: Psoriasis  
; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-09-809-003A-2

Query Match 77.8%; Score 35; DB 2; Length 9;  
Best Local Similarity 77.8%; Pred. No. 5e+05;

```
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AQYEDAIQK 9
Db 1 AQYEDIAQK 9

RESULT 3
US-09-949-016-8152
; Sequence 8152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8152
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8152

Query Match 77.8%; Score 35; DB 2; Length 637;
Best Local Similarity 77.8%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AQYEDAIQK 9
Db 356 AQYEDIAQK 364

RESULT 4
US-09-248-796A-17724
; Sequence 17724, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17724
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17724

Query Match 73.3%; Score 33; DB 2; Length 390;
Best Local Similarity 75.0%; Pred. No. 98;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QYEDIAQK 9
Db 288 QYHDSIQK 295

RESULT 5
```

```
US-09-252-991A-16954
; Sequence 16954, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16954
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16954

Query Match 73.3%; Score 33; DB 2; Length 527;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQYEDAIQ 8
Db 415 AQYEKAIQ 422

RESULT 6
US-08-725-532A-3
; Sequence 3, Application US/08725532A
; Patent No. 6020179
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,532A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0130 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
```

## US-08-725-532A-3

Query Match 71.1%; Score 32; DB 2; Length 173;  
Best Local Similarity 85.7%; Pred. No. 65;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8  
Db 125 KYEDAIQ 131

## RESULT 7

US-09-164-193-2  
; Sequence 2, Application US/09164193C  
; Patent No. 6258582

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan L.

; TITLE OF INVENTION: NOVEL CSAPTP NUCLEIC ACID MOLECULES AND USES THEREFOR (as amended)

; FILE REFERENCE: MNI-051

; CURRENT APPLICATION NUMBER: US/09/164,193C

; CURRENT FILING DATE: 1998-09-30

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 173

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-164-193-2

Query Match 71.1%; Score 32; DB 2; Length 173;  
Best Local Similarity 85.7%; Pred. No. 65;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8  
Db 125 KYEDAIQ 131

## RESULT 8

US-09-221-448A-2

; Sequence 2, Application US/09221448A

; Patent No. 6436685

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPTP PROTEIN MOLECULES AND USES THEREFOR (AS AMENDED)

; FILE REFERENCE: MNI-051DVI

; CURRENT APPLICATION NUMBER: US/09/221,448A

; CURRENT FILING DATE: 1998-12-28

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 173

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-221-448A-2

Query Match 71.1%; Score 32; DB 2; Length 173;  
Best Local Similarity 85.7%; Pred. No. 65;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDAIQ 8  
Db 125 KYEDAIQ 131

## RESULT 9

US-10-163-214-10

; Sequence 10, Application US/10163214

; Patent No. 6849781

; GENERAL INFORMATION:

; APPLICANT: Allen, Stephen M.

; APPLICANT: Broglie, Karen E.

; APPLICANT: Butler, Karlene H.

; APPLICANT: Thorpe, Catherine J.  
; TITLE OF INVENTION: Starch Synthase Isoform V  
; FILE REFERENCE: BB1520 US NA  
; CURRENT APPLICATION NUMBER: US/10/163,214  
; CURRENT FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: 60/297,099  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 10  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-163-214-10

Query Match 71.1%; Score 32; DB 2; Length 293;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9  
Db 276 SQVEDIYQK 284

## RESULT 10

US-09-543-681A-6875

; Sequence 6875, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETTON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 6875

; LENGTH: 315

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-6875

Query Match 71.1%; Score 32; DB 2; Length 315;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQVEDAIQK 9  
Db 105 AQVEDKAQR 113

## RESULT 11

US-09-889-463A-38

; Sequence 38, Application US/09889463A

; Patent No. 6680185

; GENERAL INFORMATION:

; APPLICANT: Cahoon, Rebecca E.

; APPLICANT: Falco, Saverio C.

; APPLICANT: Kinney, Anthony J.

; APPLICANT: Miao, Guo-Hua

; TITLE OF INVENTION: Plant Polyphenol Oxidase Homologs

; FILE REFERENCE: BB1330

; CURRENT APPLICATION NUMBER: US/09/889,463A

; CURRENT FILING DATE: 2001-07-16

; PRIOR APPLICATION NUMBER: 60/119,590

; PRIOR FILING DATE: 1999-02-10

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 38

; LENGTH: 358

; TYPE: PRT

```
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (5)
; OTHER INFORMATION: Xaa = any amino acid
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (20)
; OTHER INFORMATION: Xaa = any amino acid
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (179)
; OTHER INFORMATION: Xaa = any amino acid
US-09-889-463A-38

Query Match      71.1%; Score 32; DB 2; Length 358;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIO 8
Db 43 AKYDDAVQ 50

RESULT 12
US-09-328-352-6614
; Sequence 6614, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6614
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6614

Query Match      71.1%; Score 32; DB 2; Length 424;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YEDAIOK 9
Db 194 YEDAIOK 200

RESULT 13
US-09-689-065B-6
; Sequence 6, Application US/09689065B
; Patent No. 6605696
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products, Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS AND RELATED METHODS AND
; FILE REFERENCE: 3153.00187/PC10589A
; CURRENT APPLICATION NUMBER: US/09/689,065B
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US Prov. 60/160,922
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US Prov. 60/163,858
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-09-689-065B-6
```

```
Query Match      71.1%; Score 32; DB 2; Length 812;
Best Local Similarity 55.6%; Pred. No. 3.5e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIOK 9
Db 247 AEYEEALQE 255

RESULT 14
US-10-210-296A-6
; Sequence 6, Application US/10210296A
; Patent No. 6982114
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products, Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED METHODS AND
; FILE REFERENCE: 3153.00355/PC10589B
; CURRENT APPLICATION NUMBER: US/10/210,296A
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US 09/689,065
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/160,922
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US 60/163,858
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-10-210-296A-6

Query Match      71.1%; Score 32; DB 2; Length 812;
Best Local Similarity 55.6%; Pred. No. 3.5e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDAIOK 9
Db 247 AEYEEALQE 255
```

```
RESULT 15
US-10-163-214-12
; Sequence 12, Application US/10163214
; Patent No. 6849781
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Broglie, Karen E.
; APPLICANT: Butler, Karlene H.
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Starch Synthase Isoform V
; FILE REFERENCE: BB1520 US NA
; CURRENT APPLICATION NUMBER: US/10/163,214
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,099
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-163-214-12

Query Match      71.1%; Score 32; DB 2; Length 914;
Best Local Similarity 66.7%; Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQYEDAIOK 9
Db 247 AEYEEALQE 255
```

Db 897 SQYEDIYQK 905

Search completed: July 5, 2006, 19:24:24  
Job time : 21.2718 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:20:41 ; Search time 6.41667 Seconds  
(without alignments)  
104.964 Million cell updates/sec

Title: US-10-774-928A-5  
Perfect score: 33  
Sequence: 1 AEASLY 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	553	2 I59009	epidermal keratin
2	33	100.0	643	1 KRH02	keratin 1, type II
3	31	93.9	483	2 A34720	keratin 8, type II
4	30	90.9	243	2 JC5723	capsular polysacch
5	30	90.9	249	2 B87309	hypothetical prote
6	30	90.9	355	2 T47147	hypothetical prote
7	30	90.9	423	2 S54135	phytoene synthase
8	30	90.9	534	2 I37942	keratin 4, type II
9	30	90.9	602	2 AF2771	single-stranded-DN
10	30	90.9	602	2 AF2771	single-stranded-DN
11	30	90.9	629	2 A29666	keratin, 65K type
12	30	90.9	710	2 AG2285	hypothetical prote
13	29	87.9	268	2 C83104	hypothetical prote
14	29	87.9	351	2 T26940	hypothetical prote
15	29	87.9	364	2 T47698	mannose-1-phosphat
16	29	87.9	386	2 T00589	hookless1-like pro
17	29	87.9	444	2 E96905	TPR-repeat-contain
18	29	87.9	494	2 A10964	probable carboxydr
19	29	87.9	585	2 T15995	hypothetical prote
20	29	87.9	591	2 A84765	hypothetical prote
21	29	87.9	629	2 T45829	hypothetical prote
22	29	87.9	669	2 A97443	hypothetical prote
23	29	87.9	669	2 AC2661	hypothetical prote
24	29	87.9	815	2 T19704	hypothetical prote
25	29	87.9	1758	2 T34393	hypothetical prote
26	28	84.8	216	2 JS0639	GTP-binding protei
27	28	84.8	217	2 S38741	GTP-binding protei
28	28	84.8	217	2 S45023	GTP-binding protei
29	28	84.8	217	2 S41431	GTP-binding protei

GTP-binding protei  
cytokeratin 8, typ  
keratin, 53K type  
keratin 8, type II  
keratin 8, type II  
cytokeratin Endoa  
hypothetical prote  
hypothetical prote  
OmpA-related prote  
GTP-binding protei  
ribonuclease PH (i  
keratin-like prote  
sulfate adenylate  
3-hydroxyisobutyra  
keratin, type II -  
keratin 6d, type I

28 84.8 218 2 T06448  
28 84.8 310 2 S43865  
28 84.8 370 2 A25004  
28 84.8 487 2 JT0407  
28 84.8 489 2 S05474  
28 84.8 490 2 JS0658  
28 84.8 627 2 T25395  
28 84.8 646 2 T34532  
28 84.8 1055 2 A87364  
28 84.8 214 2 T04872  
27 81.8 238 2 G87267  
27 81.8 257 2 I38025  
27 81.8 266 2 C96913  
27 81.8 298 2 C87417  
27 81.8 346 2 S34165  
27 81.8 384 2 I61769

## ALIGNMENTS

### RESULT 1

I59009  
epidermal keratin subunit II - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: I59009; A61205  
R;Steinert, P.M.; Parry, D.A.D.; Racoosin, E.L.; Idler, W.W.; Steven, A.C.; Trus, B.L.; Tanooi  
Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713, 1984  
A;Title: The complete cDNA and deduced amino acid sequence of a type II mouse epidermal  
A;Reference number: I59009; MUID:85014838; PMID:6207530  
A;Accession: I59009  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-553 <RES>  
A;Cross-references: UNIPROT:P50446; UNIPARC:UPI000016CE87; GB:K02108; NID:g198634; PTDN:  
R;Finch, J.; Andrews, K.; Krieg, P.; Fuerstenberger, G.; Slaga, T.; Ootsuyama, A.; Tanooi  
Carcinogenesis 12, 1519-1522, 1991  
A;Title: Identification of a cloned sequence activated during multi-stage carcinogenesis  
A;Reference number: A61205; MUID:91316763; PMID:1713533  
A;Accession: A61205  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 528-553 <FIN>  
A;Cross-references: UNIPARC:UPI00001774C8  
C;Genetics:  
A;Gene: KER2  
C;Superfamily: cytoskeletal keratin

Query Match 100.0%; Score 33; DB 2; Length 553;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7

Db 339 AEASLY 345

### RESULT 2

KRH02  
keratin 1, type II, cytoskeletal - human  
N;Alternate names: 67K type II epidermal keratin; cytokeratin 1  
C;Species: Homo sapiens (man)  
C;Date: 04-Dec-1986 #sequence\_revision 22-Oct-1999 #text\_change 10-Dec-1999  
C;Accession: A22940; A02950; A43342  
R;Johnson, L.D.; Idler, W.W.; Zhou, X.M.; Roop, D.R.; Steinert, P.M.  
Proc. Natl. Acad. Sci. U.S.A. 82, 1896-1900, 1985  
A;Reference number: A22940; MUID:85166239; PMID:2580302  
A;Accession: A22940  
A;Molecule type: DNA  
A;Residues: 1-643 <JOH>  
A;Cross-references: UNIPARC:UPI0000173D59; GB:M98776; GB:M11215; GB:M11845; GB:M11846; NJ

A;Note: translation of initiator Met is not shown





Amino acids: 463/27 475/1  
C; Superfamily: cytoskeletal keratin  
C; Keywords: coiled coil; intermediate filament  
F; 1-152/Domain: head #status predicted <HEA>  
F; 153-464/Domain: helical rod #status predicted <ROD>

F/465-534/Domain: tail #status predicted <TAI>

Query Match 90.9%; Score 30; DB 2; Length 534;

Best Local Similarity 85.7%; Pred. No. 49; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7

|||||

Db 338 AEAEALY 344

RESULT 9

E97551 single-stranded-DNA-specific exonuclease recJ (recJ) rp528 [imported] - Agrobacterium tumefaciens

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004

C:Accession: E97551

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: E97551

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-602 <KUR>

A:Cross-references: UNIPROT:Q8UF16; UNIPARC:UPI00000D1BE7; GB:AE007869; PIDN:AAK87366.1;

C:Genetics: AGR\_C\_2915

A:Gene: AGR\_C\_2915

A:Map position: circular chromosome

C:Superfamily: single-stranded-DNA-specific exonuclease RecJ

Query Match

Best Local Similarity 90.9%; Score 30; DB 2; Length 602;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7

|||||

Db 363 AEAEALY 369

RESULT 10

AF2771 single-stranded-DNA-specific exonuclease [imported] - Agrobacterium tumefaciens (strain AF2771)

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004

C:Accession: AF2771

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AF2771

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-602 <KUR>

A:Cross-references: UNIPROT:Q8UF16; UNIPARC:UPI00000D1BE7; GB:AE008688; PIDN:AAL42588.1;

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: recJ

A:Map position: circular chromosome

C:Superfamily: single-stranded-DNA-specific exonuclease RecJ

Query Match

Best Local Similarity 90.9%; Score 30; DB 2; Length 602;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7

|||||

Db 363 AEAEALY 369

RESULT 11

A29666

keratin, 65K type II cytoskeletal - human

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 20-Apr-2001

C:Accession: A29666

R:Klinge, E.M.; Sylvestre, Y.R.; Freedberg, I.M.; Blumenberg, M.

J. Mol. Evol. 24, 319-329, 1987

A:Title: Evolution of keratin genes: different protein domains evolve by different pathways

A:Reference number: A29666; MUID:87254239; PMID:2499698

A:Accession: A29666

A:Molecule type: DNA

A:Residues: 1-629 <KLI>

A:Cross-references: UNIPARC:UPI00000161C35; GB:X05418; NID:934040; PIDN:CAA28991.1; PID:G92;

92; PIDN:CAA28994.1; PID:G1335193; PIDN:CAA28995.1; PID:G1335194; GB:X03421; NID:G34047;

A>Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 21

A>Note: the complete translation is not annotated in GenBank entries HSKER65A, HSKER65B,

s mistranslated as a Met initiator codon

C:Genetics:

A:Introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match

Best Local Similarity 90.9%; Score 30; DB 2; Length 629;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7

|||||

Db 387 AEAEALY 393

RESULT 12

AG2285

hypothetical protein all3838 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C:Accession: AG2285

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AG2285

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-710 <KUR>

A:Cross-references: UNIPROT:Q8Y015; UNIPARC:UPI00000CE918; GB:BA000019; PIDN:BA075537.1;

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all3838

Query Match

Best Local Similarity 90.9%; Score 30; DB 2; Length 710;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7

|||||

Db 500 SEAESLY 506

RESULT 13

C83104

hypothetical protein PA4327 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004

C:Accession: C83104

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Brj

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: C83104  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-268 <STO>  
A;Cross-references: UNIPROT:Q9HW74; UNIPARC:UPI0000005CBF; GB:AE004849; GB:AE004091; NID  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA4327

Query Match 87.9%; Score 29; DB 2; Length 268;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEAESELY 7  
|||  
Db 70 AEAEKLY 76

## RESULT 14

T26840  
hypothetical protein Y43F4B.2 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T26840  
R;Matthews, L.  
submitted to the EMBL Data Library, January 1998  
A;Reference number: Z20276  
A;Accession: T26840  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-351 <WIL>  
A;Cross-references: UNIPROT:O45931; UNIPARC:UPI000017BC75; EMBL:AL021481; PIDN:CAA16331  
A;Experimental source: clone Y43F4B  
C;Genetics:  
A;Gene: CESP:Y43F4B.2  
A;Introns: 38/3; 125/3; 150/1; 200/1; 248/1; 274/3

Query Match 87.9%; Score 29; DB 2; Length 351;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EAESLY 7  
|||  
Db 93 EAESLY 98

## RESULT 15

T47698  
mannose-1-phosphate guanylyltransferase-like protein - *Arabidopsis thaliana*  
N;Alternate names: protein T22E16.250  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C;Accession: T47698  
R;Renes, V.; Wurmbach, E.; Drzonek, H.; Ansoerge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.H.  
submitted to the Protein Sequence Database, February 2000  
A;Reference number: Z24472  
A;Accession: T47698  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-364 <BEN>  
A;Cross-references: UNIPROT:Q9M2S0; UNIPARC:UPI000000A4F69; EMBL:AL132975  
A;Experimental source: cultivar Columbia; BAC clone T22E16  
C;Genetics:  
A;Map position: 3  
A;Introns: 37/3; 58/3; 141/3  
A;Note: T22E16.250  
C;Superfamily: mannose-1-phosphate guanylyltransferase

Query Match 87.9%; Score 29; DB 2; Length 364;  
Best Local Similarity 85.7%; Pred. No. 57;

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:01:48 ; Search time 81.375 Seconds  
(without alignments)  
79.571 Million cell updates/sec

Title: US-10-774-928A-5  
Perfect score: 33  
Sequence: 1 AEASLY 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 92501592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	221	Q94J09_ORYSA	Q2y3h7 oryza sativ
2	33	100.0	311	Q2Y5H7_NITMU	Q2y3h7 nitrosospi
3	33	100.0	330	Q4K8B8_PSEF5	Q4k8b8 pseudomonas
4	33	100.0	340	Q2S2S3_RHIME	Q92sz3 rhizobium m
5	33	100.0	643	K2C1_HUMAN	P04264 homo sapien
6	33	100.0	771	Q3VPN4_9CHLB	Q3vpn4 pelodictyon
7	33	100.0	5303	Q9V628_DROME	Q9v628 drosophila
8	31	93.9	98	Q4Z999_9CAUD	Q4z999 staphylococ
9	31	93.9	220	Q75119_ORYSA	Q75119 oryza sativ
10	31	93.9	279	Q7L4M3_HUMAN	Q7l4m3 homo sapien
11	31	93.9	365	Q96910_HUMAN	Q96910 homo sapien
12	31	93.9	482	K2C8_HUMAN	P05787 homo sapien
13	31	93.9	483	Q53GJ0_HUMAN	Q53gj0 homo sapien
14	31	93.9	483	Q6DHW5_HUMAN	Q6dhw5 homo sapien
15	31	93.9	483	Q6GMY0_HUMAN	Q6gmy0 homo sapien
16	31	93.9	483	Q5RB34_PONPY	Q5rb34 pongo pygma
17	31	93.9	499	Q9H552_HUMAN	Q9h552 homo sapien
18	31	93.9	505	Q6P4C7_HUMAN	Q6p4c7 homo sapien
19	31	93.9	667	Q35LE8_9GAMM	Q35le8 alkalilimni
20	30	90.9	74	Q9Y957_NPVEP	Q9y957 epiphyas po
21	30	90.9	120	Q53WA3_THET8	Q53wa3 thermus the
22	30	90.9	120	Q746N7_THET8	Q746n7 thermus the
23	30	90.9	147	Q4RD55_MACFA	Q4rd55 macaca fasc
24	30	90.9	151	Q510X7_HUMAN	Q510x7 homo sapien
25	30	90.9	157	Q46DA8_METEA	Q46da8 methanosarc
26	30	90.9	176	Q44SN2_SOLUS	Q44sn2 solibacter
27	30	90.9	183	Q8GPF5_PSEAE	Q8gpf5 pseudomonas
28	30	90.9	184	Q5CHF8_CRYHO	Q5chf8 cryptospori
29	30	90.9	184	Q5CTQ9_CRYPV	Q5ctq9 cryptospori
30	30	90.9	243	Q87929_STRTR	Q87929 streptococc
31	30	90.9	243	Q55203_STRSL	Q55203 streptococc

32	30	90.9	243	2	Q56039_STRTR	Q56039 streptococc
33	30	90.9	243	2	Q7B099_STRTR	Q7b099 streptococc
34	30	90.9	243	2	Q8G954_STRTR	Q8g954 streptococc
35	30	90.9	243	2	Q8GMA7_STRTR	Q8gma7 streptococc
36	30	90.9	243	2	Q8GP56_STRTR	Q8gp56 streptococc
37	30	90.9	243	2	Q8GP78_STRTR	Q8gp78 streptococc
38	30	90.9	243	2	Q8GP94_STRTR	Q8gp94 streptococc
39	30	90.9	243	2	Q8GPD8_STRTR	Q8gpd8 streptococc
40	30	90.9	243	2	Q5LZN2_STRT1	Q5lzn2 streptococc
41	30	90.9	243	2	Q5M488_STRT2	Q5m488 streptococc
42	30	90.9	249	2	Q9AAW0_CAUCR	Q9aaw0 caulobacter
43	30	90.9	254	2	Q4ZS06_PSEU2	Q4zsq6 pseudomonas
44	30	90.9	290	2	Q882J0_PSESM	Q882j0 pseudomonas
45	30	90.9	319	2	Q36NC9_MARHY	Q36nc8 marinobacte

## ALIGNMENTS

RESULT 1  
Q94J09\_ORYSA PRELIMINARY; PRT; 221 AA.  
ID Q94J09; Q7XJ38;  
AC Q94J09; integrated into UniProtKB/TrEMBL.  
DT 01-DEC-2001, sequence version 1.  
DT 21-FEB-2006, entry version 24.  
DE GTP-binding protein GTP1.  
GN Name=P0481E1.49-1;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP Clade;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22337376; PubMed=12447438; DOI=10.1038/nature01184;  
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,  
RA Wu J., Nilmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,  
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,  
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,  
RA Ikono M., Ito S., Ito Y., Ito Y., Iwabuchi A., Kamiya K.,  
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,  
RA Machita K., Machara T., Mizuno H., Mizubayashi T., Mukai Y.,  
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,  
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,  
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,  
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,  
RA Zhong H., Iwana H., Endo T., Ito H., Hahn J.H., Kim H.-I., Eun M.-Y.,  
RA Yano M., Jiang J., Gojobori T.;  
RT "The genome sequence and structure of rice chromosome 1.";  
RL Nature 420:312-316(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RA Quanhong Y., Rihe P., Aisheng X.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Attached to the cytoplasmic side of the  
CC membrane by a lipid-anchor (By similarity).  
CC -!- SIMILARITY: Belongs to the small GTPase superfamily. Rab family.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs license  
DR EMBL; AP003076; BAB56054.1; -; Genomic\_DNA.  
DR EMBL; AX327044; AAP92129.1; -; mRNA.  
DR HSSP; P07560; 1G16.  
DR SMR; Q94J09; 13-180.  
DR Gramene; Q94J09; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005525; F:GTP binding; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0015031; P:protein transport; IEA.  
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.

```

DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR003577; GTPase_Rab.
DR InterPro; IPR003578; GTPase_Rho.
DR InterPro; IPR002041; RAN_GTP_bd.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR005225; Small_GTP_bd.
DR Pfam; PF00071; Ras; 1.
DR PRINTS; PR00449; RASTRNSFRNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR GTP-binding; Membrane; Nucleotide-binding.
KW GTP-binding; Membrane; Nucleotide-binding.
SQ SEQUENCE 221 AA; 24193 MW; 075EB2A613B7B7F1 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 221;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
Db 150 AEAESLY 156

RESULT 2
Q2Y5H7 NITMU
ID Q2Y5H7 NITMU PRELIMINARY; PRT; 311 AA.
AC Q2Y5H7
DT 20-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Kinesin light chain precursor.
GN ORFNames=Nmul_A2707;
OS Nitrosospora multiformis (strain ATCC 25196 / NCIMB 11849).
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadales; Nitrosospora.
OX NCBI_TaxID=323848;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25196;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Chain P., Malfatti S., Shin M.,
RA Vergez L., Schmutz J., Larimer F., Land M., Hauser L., Kyrpides N.,
RA Lykidis A., Richardson P.;
RT "Complete sequence of Chromosome 1 of Nitrosospora multiformis ATCC
RT 25196";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000103; ABB75994.1; -; Genomic_DNA.
KW SIGNAL.
FT SIGNAL. 1 22 Potential.
SQ SEQUENCE 311 AA; 34152 MW; AB10459D7863E758 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 311;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
Db 126 AEAESLY 132

RESULT 3
Q4K8B8_PSEF5
ID Q4K8B8_PSEF5 PRELIMINARY; PRT; 330 AA.
AC Q4K8B8;
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, entry version 1.
DE Cobalamin biosynthesis protein CobC.
GN OrderedLocusNames=PFL_4427;

```

```

OS Pseudomonas fluorescens (strain Pf-5 / ATCC BAA-477).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=220664;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15980861; DOI=10.1038/nbt1110;
RA Paulsen I.T., Press C.M., Ravel J., Kobayashi D.Y., Myers G.S.A., Dodson R.J.,
RA Mavrodi D.V., Deboy R.T., Seshadri R., Ren Q., Madupu R., Sullivan S.A.,
RA Durkin A.S., Brinkac L.M., Daugherty S.C., Sullivan S.A.,
RA Rosovitz M.J., Winn M.L., Zhou L., Schneider D.J., Cartinhour S.W.,
RA Nelson W.C., Weidman J., Watkins K., Tran K., Khouri H., Pierson E.A.,
RA Pierson L.S. III, Thomashow L.S., Loper J.E.;
RT "Complete genome sequence of the plant commensal Pseudomonas
RT fluorescens Pf-5.";
RT Nat. Biotechnol. 23:873-878 (2005).
RL Nat. Biotechnol. 23:873-878 (2005).
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasm (By similarity).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000076; AAY93678.1; -; Genomic DNA.
DR GO; GO:0016769; P:transferase activity; transferring nitrogen. .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR004838; NHtransf_1_BS.
DR Pfam; PF00155; Aminotran_1_2; 1.
DR PROSITE; PS00105; AA_TRANSFER_CLASS_1; UNKNOW_1.
KW Complete proteome; Pyridoxal phosphate.
SQ SEQUENCE 330 AA; 36974 MW; EEC4E27BF648D117 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 330;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
Db 279 AEAESLY 285

RESULT 4
Q92S23 RHIME
ID Q92S23 RHIME PRELIMINARY; PRT; 340 AA.
AC Q92S23
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.
DT 01-DEC-2001, sequence version 1.
DE Hypothetical protein.
GN OrderedLocusNames=R00202; ORFNames=SMC02877;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampeger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AL591782; CAC41589.1; -; Genomic DNA.
DR BioCyc; SMEL382.SMC02877-MONOMER; -.
DR GO; GO:0005529; F:sugar binding; IEA.

```

DR GO: 0005975; P: carbohydrate metabolism; IEA.  
DR InterPro: IPR001347; SIS.  
DR Pfam: PF01380; SIS; 3.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 340 AA; 35192 MW; A6DBB8114F602DF7 CRC64;  
Query Match 100.0%; Score 33; DB 2; Length 340;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AEASLSY 7  
Db 199 AEASLSY 205  
|||||  
RESULT 5  
ID K2C1\_HUMAN STANDARD; PRT; 643 AA.  
AC P04264; Q14720; Q6GSJ0; Q9H298;  
DT 20-MAR-1987, integrated into UniProtKB/Swiss-Prot.  
DT 01-FEB-1996, sequence version 4.  
DT 07-FEB-2006, entry version 76.  
DE Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1)  
DE (K1) (67 kDa cytokeatin) (Hair alpha protein).  
GN Name=KRT1; Synonyms=KRTA;  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RX MEDLINE=85166239; PubMed=2580302;  
RA Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;  
RT "Structure of a gene for the human epidermal 67-kDa keratin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RA Whitlock N.V., Eady R.A.J., McGrath J.A.;  
RT "Genomic organization of the human keratin 1 gene.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RA Hattell S.J., Eady R.A.J., Wennerstrand L., Dopping-Hepenstal P.J.,  
RA Leigh I.M., Munro C., Kelsell D.P.;  
RT "Novel splice site mutation in keratin 1 underlies mild epidermolytic  
RT palmoplantar keratoderma in three kindreds.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA], AND VARIANT TYR-357.  
RC TISSUE=Skin;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Abrahams R.D., Mullahy S.J.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP PRELIMINARY NUCLEOTIDE SEQUENCE [MRNA] OF 151-643.  
RX MEDLINE=85207740; PubMed=2581964;  
RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,  
RA Roop D.R.;  
RT "Amino acid sequences of mouse and human epidermal type II keratins of  
RT Mr 67,000 provide a systematic basis for the structural and functional  
RT diversity of the end domains of keratin intermediate filament  
RT subunits.";  
RL J. Biol. Chem. 260:7142-7149(1985).  
RN [6]  
RP PROTEIN SEQUENCE OF 1-7.  
RC TISSUE=Platelet;  
RX MEDLINE=22608298; PubMed=12665801; DOI=10.1038/nbt810;  
RA Gevaert K., Goethals W., Martens L., Van Damme J., Staes A.,  
RA Thomas G.R., Vandekerckhove J.;  
RT "Exploring proteomes and analyzing protein processing by mass  
RT spectrometric identification of sorted N-terminal peptides.";  
RL Nat. Biotechnol. 21:566-569(2003).  
RN [7]  
RP CITRULLINATION.  
RX MEDLINE=96374388; PubMed=8780679; DOI=10.1006/bbrc.1996.1240;  
RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;  
RT "Preferential demethylation of keratin K1 and filaggrin during the  
RT terminal differentiation of human epidermis.";  
RL Biochem. Biophys. Res. Commun. 225:712-719(1996).  
RN [8]  
RP VARIANT EHK PRO-160.  
RX MEDLINE=92386601; PubMed=1381288; DOI=10.1016/0092-8674(92)90315-4;  
RA Chipev C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,  
RA Compton J.G., Steinert P.M.;  
RT "A leucine-->proline mutation in the H1 subdomain of keratin 1 causes  
RT epidermolytic hyperkeratosis.";  
RL Cell 70:821-828(1992).  
RN [9]  
RP VARIANT ALLELE 1B 559-GLY--TYR-565 DEL.  
RX MEDLINE=93107743; PubMed=1281859; DOI=10.1111/1523-1747.ep12614149;  
RA Korge B.P., Compton J.G., Steinert P.M., Mischke D.;  
RT "The two size alleles of human keratin 1 are due to a deletion in the  
RT glycine-rich carboxyl-terminal V2 subdomain.";  
RL J. Invest. Dermatol. 99:697-702(1992).  
RN [10]  
RP VARIANT EHK GLN-489.  
RX MEDLINE=92376531; PubMed=1380725;  
RA Rothnagel J.A., Dorney A.M., Dempsey L.D., Longley M.A.,  
RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;  
RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic  
RT hyperkeratosis.";  
RL Science 257:1128-1130(1992).  
RN [11]  
RP VARIANT EHK CYS-481.  
RX MEDLINE=94216497; PubMed=7512983;  
RA Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;  
RT "Genetic mutations in the K1 and K10 genes of patients with  
RT epidermolytic hyperkeratosis. Correlation between location and disease  
RT severity.";  
RL J. Clin. Invest. 93:1533-1542(1994).  
RN [12]  
RP VARIANTS EHK GLY-154; SER-187 AND PRO-192.  
RX MEDLINE=94117869; PubMed=7507151; DOI=10.1111/1523-1747.ep12371725;  
RA Yang J.-M., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,  
RA Steinert P.M., Compton J.G.;  
RT "Mutations in the H1 and IA domains in the keratin 1 gene in  
RT epidermolytic hyperkeratosis.";  
RL J. Invest. Dermatol. 102:17-23(1994).  
RN [13]  
RP VARIANTS EHK PRO-185 AND SER-187.  
RX MEDLINE=94117870; PubMed=7507152; DOI=10.1111/1523-1747.ep12371726;  
RA McLean W.H.I., Bady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,  
RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,  
RA Morley S.M.;  
RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous  
RT congenital ichthyosiform erythroderma (BCIE).";  
RL J. Invest. Dermatol. 102:24-30(1994).

[14]  
RP VARIANT NEPPK IL8-73.  
RX MEDLINE=95096501; PubMed=7528239; DOI=10.1111/1523-1747.epi12412771;  
RA Kimonis V., DiGiovanna J.J., Yang J.-M., Doyle S.Z., Bale S.J.,  
RA Compton J.G.;  
RT "A mutation in the VI end domain of keratin 1 in non-epidermolytic  
RT palmar-plantar keratoderma";  
RL J. Invest. Dermatol. 103:764-769(1994).  
RN [15]  
RP VARIANT EHK VAL-339.  
RX MEDLINE=99072666; PubMed=9856846;  
RX DOI=10.1046/j.1523-1747.1998.00382.x;  
RA Kremer H., Lavrijsen A.P., McLean W.H.I., Lane E.B., Melchers D.,  
RA Ruiter D.J., Mariman A.C., Steijlen P.M.;  
RT "An atypical form of bullous congenital ichthyosiform erythroderma is  
RT caused by a mutation in the L12 linker region of keratin 1";  
RL J. Invest. Dermatol. 111:1224-1226(1998).  
RN [16]  
RP VARIANTS AEI PHE-478 AND THR-478.  
RX MEDLINE=99162195; PubMed=10053007;  
RA Sybert V.P., Francis J.S., Corden L.D., Smith L.T., Weaver M.,  
RA Stephens K., McLean W.H.I.;  
RT "Cyclic ichthyosis with epidermolytic hyperkeratosis: a phenotype  
RT conferred by mutations in the 2B domain of keratin K1";  
RL Am. J. Hum. Genet. 64:732-738(1999).  
RN [17]  
RP VARIANT EHK THR-187.  
RX MEDLINE=99247363; PubMed=10232403;  
RA Arin M.J., Longley M.A., Kuster W., Huber M., Hohl D., Rothnagel J.A.,  
RA Roop D.R.;  
RT "An asparagine to threonine substitution in the 1A domain of keratin  
RT 1: a novel mutation that causes epidermolytic hyperkeratosis";  
RL Exp. Dermatol. 8:124-127(1999).  
RN [18]  
RP VARIANT AEI PHE-478.  
RX MEDLINE=20062276; PubMed=10597140;  
RA Michael E.J., Schneiderman P., Grossman M.E., Cristiano A.M.;  
RT "Epidermolytic hyperkeratosis with polycyclic psoriasiform plaques  
RT resulting from a mutation in the keratin 1 gene";  
RL Exp. Dermatol. 8:501-503(1999).  
RN [19]  
RP VARIANT EHK PRO-213.  
RX MEDLINE=20305470; PubMed=10844506;  
RX DOI=10.1046/j.1365-2230.2000.00625.x;  
RA Cserhalmi-Friedman P.B., Squeo R., Gordon D., Garzon M.,  
RA Schneiderman P., Grossman M.E., Cristiano A.M.;  
RT "Epidermolytic hyperkeratosis in a Hispanic family resulting from a  
RT mutation in the keratin 1 gene";  
RL Clin. Exp. Dermatol. 25:241-243(2000).  
RN [20]  
RP VARIANT EHK THR-478.  
RX MEDLINE=20151078; PubMed=10688370;  
RX DOI=10.1034/j.1600-0625.2000.00901016.x;  
RA Arin M.J., Longley M.A., Epstein E.H. Jr., Rothnagel J.A., Roop D.R.;  
RT "Identification of a novel mutation in keratin 1 in a family with  
RT epidermolytic hyperkeratosis";  
RL Exp. Dermatol. 9:116-119(2000).  
RN [21]  
RP VARIANT EHK ASP-154.  
RX MEDLINE=21423058; PubMed=11531804;  
RX DOI=10.1046/j.1365-2133.2001.04327.x;  
RA Whittock N.V., Ashton G.H.S., Griffiths W.A.D., Eady R.A.J.,  
RA McGrath J.A.;  
RT "New mutations in keratin 1 that cause bullous congenital  
RT ichthyosiform erythroderma and keratin 2e that cause ichthyosis  
RT bullosa of Siemens";  
RL Br. J. Dermatol. 145:330-335(2001).  
RN [22]  
RP INVOLVEMENT IN IHCM.  
RX MEDLINE=21184521; PubMed=11286616;  
RX DOI=10.1046/j.1523-1747.2001.01292.x;

Query Match

100.0%; Score 33; DB 1; Length 643;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QV 1 AEASLY 7  
Db 366 AEASLY 372  
RESULT 6  
Q3VFN4\_9CHLB  
ID Q3VFN4\_9CHLB PRELIMINARY; PRT; 771 AA.  
AC Q3VFN4;  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 11-OCT-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE TPR repeat precursor.  
GN ORFNames=PphADRAFT\_0572;  
OS Pelodictyon phaeoclathratiforme BU-1.  
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
OC Chlorobium/Pelodictyon group; Pelodictyon.  
OX NCBI\_TaxID=324925;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BU-1;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,  
RA Hamon N., Irani S., Pitluck S., Richardson P.;  
RT "Sequencing of the draft genome and assembly of Pelodictyon  
RT phaeoclathratiforme BU-1";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BU-1;  
RG US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.;  
RT "Annotation of the draft genome assembly of Pelodictyon  
RT phaeoclathratiforme BU-1";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC EMBL; AAIK01000002; EAN26333.1; -; Genomic\_DNA.  
DR GO; GO:0005871; C:kinesin complex; IEA.  
DR GO; GO:0005488; F:binding; IEA.  
DR GO; GO:0003777; F:microtubule motor activity; IEA.  
DR InterPro; IPR002151; Kinesin\_light.  
DR InterPro; IPR001440; TPR\_1.  
DR InterPro; IPR013105; TPR\_2.  
DR InterPro; IPR013026; TPR\_region.  
DR Pfam; PF00515; TPR\_1; 10.  
DR PRINTS; PR00381; KINESINLIGHT.  
DR SMART; SM00028; TPR; 13.  
DR PROSITE; PS00005; TPR; 15.  
DR PROSITE; PS0293; TPR\_REGION; 1.  
KW Signal.  
FT SIGNAL.  
SQ SEQUENCE 771 AA; 86933 MW; 8ECFC325793BAC93 CRC64;  
Query Match 100.0%; Score 33; DB 2; Length 771;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QV 1 AEASLY 7  
Db 292 AEASLY 298  
RESULT 7  
Q9V628\_DROME



ID Q9V628 DROME PRELIMINARY; PRT; 5303 AA.  
 AC Q9V628; Q9V627;  
 DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
 DT 21-FEB-2006, sequence version 2.  
 DE 21-FEB-2006, entry version 32.  
 GN CG13185-PA.  
 GN Name=CG13185; ORFNames=Dmel\_CG13185;  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houshou K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Liang P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 melanogaster euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatic  
 a genomics perspective."

RL Q9V628 DROME PRELIMINARY; PRT; 5303 AA.  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 systematic review."  
 RN Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RX Berkeley Drosophila Genome Project;  
 RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
 RA Hoskins R., Stapleton M., Pacle J., Park S., Svirskas R., Smith E.,  
 RA Yu C., Rubin G.;  
 RT "Drosophila melanogaster release 4 sequence."  
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RX FlyBase;  
 RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
 CC -1- INTERACTION:  
 CC Q9Y011:NRG15; NbExp=1; IntAct=EBI-93523; EBI-102665;  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AE003824; AAF58611.2; -; Genomic\_DNA.  
 CC DR FlyBase; Q9V628;  
 CC DR FlyBase; FBgn0033661; CG13185.  
 CC GO; GO:0005515; F:protein binding; IPI.  
 CC InterPro; IPR011704; AAA\_5.  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR012099; Midasin.  
 CC Pfam; PF07726; AAA\_3; 1.  
 CC Pfam; PF07728; AAA\_5; 5.  
 CC PIRSF; PIRSF010340; Midasin; 1.  
 CC SMART; SM00382; AAA; 2.  
 CC PROSITE; PS0234; VWFA; 1.  
 KW ATP-binding; Nucleotide-binding; Transport.  
 SQ SEQUENCE 5303 AA; 605199 MW; BF3D78FBAFBI6B6C CRC64;  
 Query Match 100.0%; Score 33; DB 2; Length 5303;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLY 7  
 |||||  
 DB 3378 AEAESLY 3384  
 RESULT 8  
 Q4Z999 9CAUD  
 ID Q4Z999 9CAUD PRELIMINARY; PRT; 98 AA.  
 AC Q4Z999;  
 DT 07-JUN-2005, integrated into UniProtKB/TrEMBL.  
 DT 07-JUN-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE ORF129.  
 OS Staphylococcus phage Twort.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.  
 OX NCBI\_TaxID=55510;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15788529; DOI=10.1073/pnas.0501140102;  
 RA Kwan T., Liu J., Dubow M., Gros P., Pelletier J.;  
 RT "The complete genomes and proteomes of 27 *Staphylococcus aureus*

```
RT bacteriophages.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:5174-5179(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AY954970; AAX92417.1; -; Genomic_DNA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR000119; Hist_DNA_bd_Bac.  
DR Pfam; PF00216; Bac_DNA_binding; I.  
DR ProDom; PD000945; Bac_DNABind; 1.  
SQ SEQUENCE 98 AA; 11966 MW; 8CE5171F422EFDE3 CRC64;  
  
Query Match 93.9%; Score 31; DB 2; Length 98;  
Best Local Similarity 85.7%; Pred. No. 58;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AEAESLY 7  
Db 28 AEAESIY 34  
|||||:  
[1]  
  
RESULT 9  
Q75I19 ORYSA PRELIMINARY; PRT; 220 AA.  
AC Q75I19  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, sequence version 1.  
DE Putative GTP-binding protein RIC2.  
GN Name=B1130610.13;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
OC Eriocarpaceae; Oryzaceae; Oryza.  
OX NCBI_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Chow T.-Y., Haing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,  
RA Chao Y.-T., Lee P.-F., Chang S.-J., Chen H.-C., Chen S.-K.,  
RA Chen T.-R., Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y.,  
RA Hsiao S.-H., Hsiung J.-N., Hsu C.-H., Kau P.-I., Lee M.-C., Leu H.-L.,  
RA Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,  
RA Wu H.-P., Shaw J.-F.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AC130603; AAT01316.1; -; Genomic_DNA.  
DR HSSP; P10114; 1KAO.  
DR SMR; Q75I19; 14-181.  
DR Gramene; Q75I19; -.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0005525; F:GTP binding; IEA.  
DR GO; GO:0015031; P:protein transport; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.  
DR InterPro; IPR003579; GTPase Rab.  
DR InterPro; IPR003577; GTPase Ras.  
DR InterPro; IPR003578; GTPase Rho.  
DR InterPro; IPR002041; RAN GTP bd.  
DR InterPro; IPR001806; Ras trnsfrmg.  
DR InterPro; IPR002078; S1954 interact.  
DR Pfam; PF00071; Ras; 1.  
DR PRINTS; PR00449; RASTRNSFRMG.  
DR SMART; SM00175; RAB; 1.  
DR TIGRFAMs; TIGR00231; small GTP; 1.  
DR PROSITE; PS00675; SIGMAS4_INTERACT_1; 1.  
SQ SEQUENCE 220 AA; 24146 MW; A4F728F377525145 CRC64;
```

```
Query Match 93.9%; Score 31; DB 2; Length 220;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AEAESLY 7  
Db 149 AEAESMY 155  
|||||:  
[1]  
  
RESULT 10  
Q7L4M3 HUMAN PRELIMINARY; PRT; 279 AA.  
AC Q7L4M3  
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.  
DT 24-MAY-2005, sequence version 1.  
DT 07-FEB-2006, entry version 6.  
DE KR78 protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield A.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RL "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Placenta;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC008200; AA08200.1; -; mRNA.  
DR HSSP; P08670; 1GK4.  
DR Ensembl; ENSG00000170421; Homo sapiens.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR003054; Keratin_II.  
DR PANTHER; PTHR1893.SF5; Keratin_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PROSITE; PS00226; IF; 1.  
SQ SEQUENCE 279 AA; 30820 MW; D9B2EB27FD985629 CRC64;  
  
Query Match 93.9%; Score 31; DB 2; Length 279;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AEAESLY 7  
Db 72 AEAESMY 78  
|||||:  
[1]
```

```
RESULT 11
Q96910 HUMAN PRELIMINARY; PRT; 365 AA.
ID Q96910
AC Q96910
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.
DT 24-MAY-2005, sequence version 2.
DT 07-FEB-2006, entry version 23.
DE KRT8 protein (Fragment).
GN Name=KRT8;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RG NIH MGC Project;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC011373; AAH11373.2; -; mRNA.
DR Ensembl; ENSG00000170421; Homo sapiens.
DR GO; GO:0005982; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR03054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
DR KW Intermediate filament.
FT NON TER 1
SQ SEQUENCE 365 AA; 41108 MW; 57CD83091D0635E8 CRC64;

Query Match 93.9%; Score 31; DB 2; Length 365;
Best Local Similarity 85.7%; Pred. NO. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEASY 7
Db 158 AEAEASY 164
|||||:|
```

RT "Posttranslational regulation of keratins: degradation of mouse and  
 RL human keratins 18 and 8.";  
 RN Mol. Cell. Biol. 9:1553-1565(1989).  
 [7]  
 RP NUCLEOTIDE SEQUENCE OF 204-482.  
 RX MEDLINE=87134779; PubMed=2434381;  
 RA Leube R.E., Bosch F.X., Romano V., Zimbelmann R., Hofler H.,  
 RA Franke W.W.;  
 RT "Cytokeratin expression in simple epithelia. III. Detection of mRNAs  
 RT encoding human cytokeratins nos. 8 and 18 in normal and tumor cells by  
 RT hybridization with cDNA sequences in vitro and in situ.";  
 RL Differentiation 33:69-85(1986).  
 [8]  
 RN PARTIAL PROTEIN SEQUENCE.  
 RP TISSUE-Colon carcinoma;  
 RX MEDLINE=97295306; PubMed=9150948;  
 RA Ji H., Reid G.E., Moritz R.L., Eddes J.S., Burgess A.W., Simpson R.J.;  
 RT "A two-dimensional gel database of human colon carcinoma proteins.";  
 RL Electrophoresis 18:605-613(1997).  
 [9]  
 RN INTERACTION WITH PNN.  
 RP PubMed=10809736; DOI=10.1074/jbc.275.20.14910;  
 RX Shi J., Sugrue S.P.;  
 RA "Dissection of protein linkage between keratins and pinin, a protein  
 RT with dual location at desmosome-intermediate filament complex and in  
 RT the nucleus.";  
 RL J. Biol. Chem. 275:14910-14915(2000).  
 [10]  
 RN PHOSPHORYLATION SITE SER-73.  
 RP MEDLINE=2190496; PubMed=11781324; DOI=10.1074/jbc.M11436200;  
 RX He T., Steplak A., Holmstrom T.H., Omary M.B., Eriksson J.E.;  
 RA "The intermediate filament protein keratin 8 is a novel cytoplasmic  
 RT substrate for c-Jun N-terminal kinase.";  
 RL J. Biol. Chem. 277:10767-10774(2002).  
 [11]  
 RN PHOSPHORYLATION SITE SER-73, AND MUTAGENESIS OF LEU-71 AND SER-73.  
 RP MEDLINE=21909510; PubMed=11788583; DOI=10.1074/jbc.M107623200;  
 RX Ku N.O., Azhar S., Omary M.B.;  
 RA "Keratin 8 phosphorylation by p38 kinase regulates cellular keratin  
 RT filament reorganization: modulation by a keratin 1-like disease  
 RT causing mutation.";  
 RL J. Biol. Chem. 277:10775-10782(2002).  
 [12]  
 RN INTERACTION WITH HCV CORE PROTEIN.  
 RP PubMed=15846844; DOI=10.1002/jmv.200401093;  
 RX Kang S.-M., Shin M.-J., Kim J.-H., Oh J.-W.;  
 RA "Proteomic profiling of cellular proteins interacting with the  
 RT hepatitis C virus core protein.";  
 RL Proteomics 5:2227-2237(2005).  
 [13]  
 RN VARIANTS CRYPTOGENIC CIRRHOSIS VAL-52; CYS-53 AND CYS-61, AND VARIANT  
 RP VAL-62.  
 RX PubMed=12724528; DOI=10.1073/pnas.0936165100;  
 RA Ku N.-O., Darling J.M., Krams S.M., Esquivel C.O., Keesfe E.B.,  
 RA Sibley R.K., Lee Y.M., Wright T.L., Omary M.B.;  
 RT "Keratin 8 and 18 mutations are risk factors for developing liver  
 RT disease of multiple etiologies.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:6063-6068(2003).  
 CC -!- SUBUNIT: Heterotetramer of two type I and two type II keratins.  
 CC Keratin 8 associates with keratin 18. Interacts with HCV core  
 CC protein and PNN.  
 CC -!- PTM: Phosphorylation on serine residues is enhanced during EGF  
 CC stimulation and mitosis. Ser-73 phosphorylation plays an important  
 CC role in keratin filament reorganization.  
 CC -!- DISEASE: Defects in KRT8 are a cause of cryptogenic cirrhosis  
 CC [MIM:215600].  
 CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
 CC microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II  
 CC (neutral to basic; 56-70 kDa) [K1 to K8].  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -!- DATABASE: NAME=Human Intermediate Filament Mutation Database;  
 CC WWW="http://www.interfil.org".  
 -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; M34225; AAA35748.1; -; mRNA.  
 DR EMBL; M34482; AAA35763.1; -; Genomic\_DNA.  
 DR EMBL; M74929; CAA52882.1; -; mRNA.  
 DR EMBL; U76549; AAB18966.1; -; mRNA.  
 DR EMBL; X98614; CAA67203.1; -; mRNA.  
 DR EMBL; M26512; AAA51542.1; -; mRNA.  
 DR EMBL; BC006654; AAH00654.2; ALT\_INIT; mRNA.  
 DR EMBL; X12882; CAA31376.1; -; mRNA.  
 DR EMBL; X74981; CAA52916.1; -; Genomic\_DNA.  
 DR PIR; A34720; A34720.  
 DR HSP; P08670; 1GK7.  
 DR INACT; P05787; -.  
 DR GLYCOSITEDB; P05787; -.  
 DR SWISS-2DPAGE; P05787; HUMAN.  
 DR Sienna-2DPAGE; P05787; -.  
 DR Ensembl; ENSG00000170421; Homo sapiens.  
 DR HGNC; HGNC:6446; KRT8.  
 DR MIM; 148050; Gene.  
 DR MIM; 215600; Phenotype.  
 DR LinkHub; P05787; -.  
 DR GO; GO:0005882; C:intermediate filament; NAS.  
 DR GO; GO:0005198; P:structural molecule activity; NAS.  
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; NAS.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin\_I.  
 DR InterPro; IPR003054; Keratin\_II.  
 DR PANTHER; PTHR1893:SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 DR KW Coiled coil; Direct protein sequencing; Disease mutation;  
 KW Intermediate filament; Keratin; Phosphorylation; Polymorphism.  
 FT INIT MET 0 0  
 FT CHAIN 1 482 Keratin, type II cytoskeletal 8.  
 FT REGION 1 89 /FTid=PRO\_0000063740.  
 FT REGION 90 397 Head.  
 FT REGION 90 125 Coil 1A.  
 FT REGION 126 142 Linker 1.  
 FT REGION 143 234 Coil 1B.  
 FT REGION 235 258 Linker 12.  
 FT REGION 259 397 Coil 2.  
 FT REGION 260 381 Necessary for interaction with PNN.  
 FT REGION 398 482 Tail.  
 FT SITE 341 341 Stutter.  
 FT MOD\_RES 23 23 Phosphoserine.  
 FT MOD\_RES 73 73 Phosphoserine (by MAPK).  
 FT MOD\_RES 431 431 Phosphoserine (by CAMK2 and MAPK).  
 FT VARIANT 52 52 G -> V (in cryptogenic cirrhosis).  
 FT VARIANT 53 53 Y -> C (in cryptogenic cirrhosis).  
 FT VARIANT 61 61 G -> C (in cryptogenic cirrhosis).  
 FT /FTid=VAR\_023058.  
 FT /FTid=VAR\_023059.  
 FT /FTid=VAR\_023060.  
 Query Match 93.9%; Score 31; DB 1; Length 482;  
 Best Local Similarity 85.7%; Pred. No. 3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLY 7  
 DB 275 AEAESMY 281  
 RESULT 13  
 Q53GJ0 HUMAN  
 IC Q53GJ0\_HUMAN PRELIMINARY; PRT; 483 AA.  
 AD Q53GJ0;  
 DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.

DT 24-MAY-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 6.  
 DE Keratin 8 variant (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94171032; PubMed=8125298; DOI=10.1016/0378-1119(94)90802-8;  
 RA Maruyama K., Sugano S.;  
 RT "Oligo-capping: a simple method to replace the cap structure of  
 RT eucaryotic mRNAs with oligoribonucleotides.";  
 RL Gene 138:171-174 (1994).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RX MEDLINE=98038986; PubMed=93731149; DOI=10.1016/S0378-1119(97)00411-3;  
 RA Suzuki Y., Yoshitomo K., Maruyama K., Sugano S.;  
 RT "Construction and characterization of a full length-enriched and a 5'-  
 RT end-enriched cDNA library.";  
 RL Gene 200:149-156(1997).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,  
 RA Tanaka A., Yokoyama S.;  
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; AK222941; BAD96661.1; -; mRNA.  
 DR Ensembl; ENSG00000170421; Homo sapiens.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin I.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR18893:SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Keratin.  
 FT NON TER 1 1  
 SQ SEQUENCE 483 AA, 53738 MW, D4F01FEB859C93D3 CRC64;  
 Query Match 93.9%; Score 31; DB 2; Length 483;  
 Best Local Similarity 85.7%; Pred. NO. 3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLY 7  
 DB 276 AEAESMY 282  
 |||||:  
 RESULT 14  
 Q6DHW5 HUMAN  
 ID Q6DHW5\_HUMAN PRELIMINARY; PRT; 483 AA.  
 AC Q6DHW5;  
 DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.  
 DT 10-MAY-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE Keratin 8.  
 GN Name=KRT8;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Liver;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Liver;  
 RA Strausberg R.;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; BC075839; AAH75839.1; -; mRNA.  
 DR Ensembl; ENSG00000170421; Homo sapiens.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin I.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR18893:SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Keratin.  
 SQ SEQUENCE 483 AA, 53704 MW, B0BC730B65929D37 CRC64;  
 Query Match 93.9%; Score 31; DB 2; Length 483;  
 Best Local Similarity 85.7%; Pred. NO. 3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLY 7  
 DB 276 AEAESMY 282  
 |||||:  
 RESULT 15  
 Q6GMY0 HUMAN  
 ID Q6GMY0\_HUMAN PRELIMINARY; PRT; 483 AA.  
 AC Q6GMY0;  
 DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 13.  
 DE Keratin 8.  
 GN Name=KRT8;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;

```
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Placenta;
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC073760; AAH73760.1; -; mRNA.
DR Ensembl; ENSG00000170421; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 483 AA; 53750 MW; 09C617DC2AFBDEE3 CRC64;

Query Match          93.9%; Score 31; DB 2; Length 483;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7
Db 276 AEASMY 282
```

Search completed: July 5, 2006, 19:18:17

Job time : 83.375 secs

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:44:12 ; Search time 64.6042 Seconds  
(without alignments)  
49.540 Million cell updates/sec

Title: US-10-774-928A-5

Perfect score: 33

Sequence: 1 AEAESLY 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq 8:\*

1: Genesep1980s:\*

2: Genesep1990s:\*

3: Genesep2000s:\*

4: Genesep2001s:\*

5: Genesep2002s:\*

6: Genesep2003as:\*

7: Genesep2003bs:\*

8: Genesep2004s:\*

9: Genesep2005s:\*

10: Genesep2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	7	6	Aao26385 Psoriasis
2	33	100.0	7	9	Adv04410 Leishmani
3	33	100.0	10	6	Aao26387 Psoriasis
4	33	100.0	10	9	Adv04412 Leishmani
5	33	100.0	12	8	Adi19066 Rat pepti
6	33	100.0	297	8	Adt60195 Plant pol
7	33	100.0	644	8	Adq17549 Human sol
8	33	100.0	644	9	Aea15501 Human pol
9	33	100.0	5303	4	Abb67866 Drosophil
10	31	93.9	125	4	Abg20747 Novel hum
11	31	93.9	180	8	Adn99441 Novel hum
12	31	93.9	198	8	Abm80639 Tumour-as
13	31	93.9	207	8	Adx73916 Plant ful
14	31	93.9	237	8	Adx76784 Plant ful
15	31	93.9	273	4	Abg26376 Novel hum
16	31	93.9	275	4	Abg15282 Novel hum
17	31	93.9	279	9	Aea90129 Human: cyt
18	31	93.9	279	10	Aef25706 Human cyt
19	31	93.9	398	6	Aae38086 Human cyt
20	31	93.9	428	8	Adp29327 Human sec
21	31	93.9	429	4	Abg19547 Novel hum
22	31	93.9	429	4	Abg19546 Novel hum
23	31	93.9	431	7	Adj70823 Human hea

24	31	93.9	476	8	ADR66033	Human pro
25	31	93.9	476	8	ADR66931	Human pro
26	31	93.9	477	4	AAU06112	Novel hum
27	31	93.9	480	10	Aef92273	Human cyt
28	31	93.9	481	6	ABU52604	Human NOV
29	31	93.9	482	6	AAE38082	Human cyt
30	31	93.9	482	8	ADT51418	Human ker
31	31	93.9	482	8	ADT51421	Human ker
32	31	93.9	482	8	ADT51419	Human ker
33	31	93.9	482	8	ADT51422	Human ker
34	31	93.9	482	8	ADT51414	Human ker
35	31	93.9	482	8	ADT51420	Human ker
36	31	93.9	482	8	ADT51415	Human ker
37	31	93.9	482	8	ADT51416	Human ker
38	31	93.9	482	8	ADT51405	Wild-type
39	31	93.9	482	8	ADT51412	Human ker
40	31	93.9	482	8	ADT51413	Human ker
41	31	93.9	482	8	ADT51417	Human ker
42	31	93.9	483	5	AAU84289	Human end
43	31	93.9	483	6	ABR47511	Breast ca
44	31	93.9	483	6	ABU57621	Different
45	31	93.9	483	6	ADA11040	Human CDN

#### ALIGNMENTS

RESULT 1  
AAO26385  
ID AAO26385 standard; peptide; 7 AA.  
XX  
AC AAO26385;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 5.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
(AKIV-) AKIVA LLC.  
XX  
O'daly JA;  
XX  
WPI; 2003-018763/01.  
XX  
New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
Claim 1; Page 43; 56pp; English.  
XX  
The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention

SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 6; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.le+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AEAESLY 7  
| | | | |  
Db 1 AEAESLY 7

RESULT 2  
ADV04410  
ID ADV04410 standard; peptide; 7 AA.  
XX  
AC ADV04410;  
XX  
DT 24-FEB-2005 (first entry)  
XX  
DE Leishmania immunotherapeutic peptide SEQ ID NO:5.  
XX  
KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;  
KW psoriasis.  
XX  
OS Leishmania sp.  
XX  
PN US2004241168-A1.  
XX  
PD 02-DEC-2004.  
XX  
PF 09-FEB-2004; 2004US-00774928.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
PR 17-OCT-2003; 2003US-00887892.  
XX  
PA (ODAL/) ODALY J A.  
XX  
PI Odaly JA;  
XX  
WPI; 2005-011563/01.  
XX  
Inhibiting selectively T-cell rolling in human, by administering compound interfering with cutaneous lymphocyte antigen-E selectin interaction and leukocyte function associated antigen-1 interaction.  
XX  
Claim 9; SEQ ID NO 5; 21pp; English.  
XX  
The invention relates to a novel method for selectively inhibiting T-cell rolling in a human host, comprising administering a compound that selectively interferes with the cutaneous lymphocyte antigen (CLA)-E selectin interaction and leukocyte function associated antigen (LFA)-1/intercellular adhesion molecule (ICAM) and very late antigen (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound includes an immunotherapeutic agent, which comprises a purified protein extract that is isolated by diethylaminoethyl Sephadex chromatography of a Nonidet P-40 insoluble particulate antigen fraction derived from isolated killed cells of anastigotes from one or more species of the Leishmania genus, where the particulate antigen fraction is solubilized with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied to diethylaminoethyl Sephadex and is eluted with a solution comprising 0.1 M sodium chloride, 8 M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract includes polypeptides having apparent molecular weights of 73, 80 and 82 kDa, after total reduction and alkylation. The species is L. amazonensis, L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the invention has antipsoriatic and immunostimulant activity, and inhibits T-cell rolling by interfering with CLA-E selectin interaction and LFA-1/ICAM and VLA/VCAM interactions. The method is useful for selectively inhibiting T-cell rolling in a human host, and for treating psoriasis. The present sequence represents a Leishmania peptide used in the invention.

SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 6; Length 10;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AEAESLY 7  
| | | | |  
Db 1 AEAESLY 7

RESULT 3  
AAO26387  
ID AAO26387 standard; peptide; 10 AA.  
XX  
AC AAO26387;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID NO 7.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
PA (AKIV-) AKIVA LLC.  
XX  
PI O'daly JA;  
XX  
WPI; 2003-018763/01.  
XX  
New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
Claim 1; Page 43; 56pp; English.  
XX  
The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention

SQ Sequence 10 AA;

Query Match 100.0%; Score 33; DB 6; Length 10;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AEAESLY 7  
| | | | |  
Db 1 AEAESLY 7

RESULT 4  
ADV04412  
ID ADV04412 standard; peptide; 10 AA.  
XX  
AC ADV04412;  
XX



DT	24-FEB-2005	(first entry)
XX	Leishmania immunotherapeutic peptide SEQ ID NO:7.	
DE	T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;	
XX	psoriasis.	
KW	Leishmania sp.	
XX	OS	
XX	US2004241168-A1.	
PN	02-DEC-2004.	
XX	09-FEB-2004; 2004US-00774928.	
XX	16-MAR-2001; 2001US-00809003.	
PR	17-OCT-2003; 2003US-00687892.	
XX	(ODAL/) ODALY J A.	
PA	Odaly JA;	
XX	WPI; 2005-011563/01.	
DR	Inhibiting selectively T-cell rolling in human, by administering compound	
XX	interfering with cutaneous lymphocyte antigen-E selectin interaction and	
PT	leukocyte function associated antigen-1 interaction.	
XX	Claim 18; SEQ ID NO 7; 21pp; English.	
PS	The invention relates to a novel method for selectively inhibiting T-cell	
XX	rolling in a human host, comprising administering a compound that	
CC	selectively interferes with the cutaneous lymphocyte antigen (CLA)-E	
CC	selectin interaction and leukocyte function associated antigen (LFA)-	
CC	1/intercellular adhesion molecule (ICAM) and very late antigen	
CC	(VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound	
CC	includes an immunotherapeutic agent, which comprises a purified protein	
CC	extract that is isolated by diethylaminoethyl Sephadex chromatography of	
CC	a Nonidet P-40 insoluble particulate antigen fraction derived from	
CC	isolated killed cells of amastigotes from one or more species of the	
CC	Leishmania genus, where the particulate antigen fraction is solubilized	
CC	with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied	
CC	to diethylaminoethyl Sephadex and is eluted with a solution comprising	
CC	0.1 M sodium chloride, 8 M urea and 0.025 M	
CC	Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract	
CC	includes polypeptides having apparent molecular weights of 73, 80 and 82	
CC	kDa, after total reduction and alkylation. The species is L.amazonensis,	
CC	L.venezuelensis, L.brasiliensis and/or L.chagasi. A compound of the	
CC	invention has antipsoriatic and immunostimulant activity, and inhibits T-	
CC	cell rolling by interfering with CLA-E selectin interaction and LFA-	
CC	1/ICAM and VLA/VCAM interactions. The method is useful for selectively	
CC	inhibiting T-cell rolling in a human host, and for treating psoriasis.	
CC	The present sequence represents a Leishmania peptide used in the	
CC	invention.	
XX	Sequence 10 AA;	
SQ	Query Match 100.0%; Score 33; DB 9; Length 10;	
	Best Local Similarity 100.0%; Pred. No. 2.7;	
	Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	1 AEAESLY 7	
DB	1 AEAESLY 7	
RESULT 5		
AD119066		
ID	AD119066 standard; peptide; 12 AA.	
XX	AC AD119066;	
XX	15-APR-2004 (first entry)	
DT		

PN US2004216190-A1.  
 XX 28-OCT-2004.  
 PD  
 XX 18-DEC-2003; 2003US-00739930.  
 PF  
 XX 28-APR-2003; 2003US-00424599.  
 PR  
 XX 28-APR-2003; 2003US-00425115.  
 PR  
 XX (KOVA/) KOVALIC D K.  
 PA  
 XX Kovalic DK;  
 PI  
 XX WPI; 2004-757369/74.  
 DR  
 XX New recombinant DNA constructs useful in the field of biochemistry and  
 PT genetics, and in particular for producing transgenic plants with improved  
 PT biological characteristics.  
 PT  
 XX Claim 2; SEQ ID NO 10272; 14pp; English.  
 PS  
 XX The invention relates a recombinant DNA construct comprising a  
 CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
 CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
 CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
 CC Arabidopsis, wheat and rape but the specification does not indicate which  
 CC sequences is derived from which organism. Also included is a method of  
 CC producing a plant having an improved property, comprising transforming a  
 CC plant with a recombinant DNA construct comprising a promoter region  
 CC functional in a plant cell operably joined to a polynucleotide encoding a  
 CC polypeptide associated with the property, and growing the transformed  
 CC plant. The property is selected from improving plant cold tolerance, for  
 CC manipulating growth rate in plant cells by modification of the cell cycle  
 CC pathway, for improving plant drought tolerance, for providing increased  
 CC resistance to plant disease, for galactomanan production, for production  
 CC of plant growth regulators, for improving plant heat tolerance, for  
 CC improving plant tolerance to herbicides, for increasing the rate of  
 CC homologous recombination in plants, for lignin production, for improving  
 CC plant tolerance to extreme osmotic conditions, for improving plant  
 CC tolerance to pathogens or pests, for yield improvement by modification of  
 CC photosynthesis, for modifying seed oil yield and/or content, for  
 CC modifying seed protein yield and/or content, for yield improvement by  
 CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC and for yield improvement by providing improved plant growth and  
 CC development under at least one stress condition. The polynucleotide may  
 CC also encode a plant transcription factor. The methods and compositions of  
 CC the present invention are useful in the field of biochemistry and  
 CC genetics, in particular for producing transgenic plants with improved  
 CC biological characteristics such as increased yield, improved nitrogen  
 CC flow, increasing plant tolerance to cold or heat, improving plant  
 CC tolerance to extreme osmotic and drought conditions, and improving plant  
 CC tolerance to plant pests or pathogens. They can also be used in physical  
 CC arrays of molecules, plant breeding markers, computer-based storage and  
 CC analysis systems. The present sequence is one of the 5544 plant protein  
 CC sequences of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
 XX  
 SQ Sequence 297 AA;  
 Query Match 100.0%; Score 33; DB 8; Length 297;  
 Best Local Similarity 100.0%; Pred. No. 89;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLY 7  
 Db 226 AEAESLY 232  
 |||||  
 RESULT 7  
 ADQ17549  
 ID ADQ17549 standard; protein; 644 AA.

XX ADQ17549;  
 AC  
 XX 26-AUG-2004 (first entry)  
 DT  
 XX Human soft tissue sarcoma-upregulated protein - SEQ ID 366.  
 DE  
 XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO2004048938-A2.  
 PN  
 XX 10-JUN-2004.  
 PD  
 XX 26-NOV-2003; 2003WO-US038193.  
 PF  
 XX 26-NOV-2002; 2002US-0429739P.  
 PR  
 XX (PROT-) PROTEIN DESIGN LABS INC.  
 PA  
 XX Aziz N, Ginsburg WM, Zlotnik A;  
 PI  
 XX WPI; 2004-441208/41.  
 DR  
 XX Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 PT  
 XX Example 2; SEQ ID NO 366; 210pp; English.  
 PS  
 XX The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC protein of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 CC  
 XX Sequence 644 AA;  
 SQ  
 Query Match 100.0%; Score 33; DB 8; Length 644;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLY 7  
 Db 367 AEAESLY 373  
 |||||  
 RESULT 8  
 AEA15501  
 ID AEA15501 standard; protein; 644 AA.  
 XX  
 AC AEA15501;  
 XX 28-JUL-2005 (first entry)  
 DT  
 XX Human polypeptide #120.  
 DE  
 XX Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
 KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
 KW cytostatic; neoplasm.  
 KW  
 XX Homo sapiens.  
 OS  
 XX

```

PN WO2005047534-A2.
XX
PD 26-MAY-2005.
XX
XX PF 15-OCT-2004; 2004WO-EP011599.
XX
PR 28-OCT-2003; 2003EP-00024565.
XX
PA (FARB ) BAYER HEALTHCARE AG.
XX
XX PI Wirtz R, Munnes M;
XX
XX WPI: 2005-372393/38.
XX
XX N-PSDB; AEA15423.
XX
PT Predicting a response to cancer treatment by detecting at least 2
PT markers, which are genes or genomic nucleic acid sequences that are
PT located on one chromosomal region, which is altered in malignant
PT neoplasia.
XX
PS Claim 7; SEQ ID NO 463; 464pp; English.
XX
CC The invention relates to a method of predicting response to cancer
CC treatment comprising detection of at least 2 markers, where the markers
CC are genes and fragments or genomic nucleic acid sequences that are
CC located on one chromosomal region, which is altered in malignant
CC neoplasia. The invention also relates to a method for the prediction,
CC diagnosis or prognosis of malignant neoplasia, methods for detecting
CC deregulations in malignant neoplasia and breast cancer, a method of
CC determining the phenotype of a cell or tissue, a method for identifying
CC genomic regions which are altered on the chromosomal level and encode
CC genes that are linked by function and are differentially expressed in
CC malignant neoplasia and breast cancer, methods of screening for agents
CC which regulate the activity of a polypeptide or a polynucleotide and
CC antibodies that specifically bind to a full length or partial
CC polypeptide. The method is useful for predicting response to cancer
CC treatment. The methods and compositions are useful for predicting,
CC diagnosing, prognosing, preventing or treating malignant neoplasia
CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,
CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell
CC lung cancer. This sequence represents a human polypeptide used in the
CC scope of the invention.
XX
SQ Sequence 644 AA;

Query Match 100.0%; Score 33; DB 9; Length 644;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
DB 367 AEAESLY 373
|||||||

RESULT 9
ABB67866
ID ABB67866 standard; protein; 5303 AA.
XX
AC ABB67866;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 30390.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
PD 27-SEP-2001.
XX

Query Match 100.0%; Score 33; DB 4; Length 5303;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
DB 3378 AEAESLY 3384
|||||||

RESULT 10
ABG20747
ID ABG20747 standard; protein; 125 AA.
XX
AC ABG20747;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20738.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS84934.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

```

PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 20; SEQ ID NO 51106; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 125 AA;

Query Match 93.9%; Score 31; DB 4; Length 125;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7

Db 115 AEAESMY 121

RESULT 11

ADN99441  
ID ADN99441 standard; protein; 180 AA.

AC ADN99441;

DT 29-JUL-2004 (first entry)

DE Novel human protein sequence #257.

XX anti-inflammatory; dermatological; neuroprotective; immunomodulator;  
KW antibacterial; virucide; antipsoriatic; cytostatic; gene therapy;  
KW vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;  
KW early aging; hormonal imbalance; ischemic heart disease;  
KW ulcerative colitis.

XX Homo sapiens.

XX WO2004038003-A2.

XX 06-MAY-2004.

XX 24-OCT-2003; 2003WO-US033947.

XX 25-OCT-2002; 2002US-0421061P.

PR 25-OCT-2002; 2002US-0421080P.

PR 25-OCT-2002; 2002US-0421552P.

PR 25-OCT-2002; 2002US-0421614P.

PR 30-OCT-2002; 2002US-0422177P.

PR 15-NOV-2002; 2002US-0422178P.

PR 15-NOV-2002; 2002US-0426355P.

PR 15-NOV-2002; 2002US-0426384P.

PR 15-NOV-2002; 2002US-0426394P.

PR 27-NOV-2002; 2002US-0429224P.  
PR 27-NOV-2002; 2002US-0429275P.  
PR 27-NOV-2002; 2002US-0429302P.  
PR 27-NOV-2002; 2002US-0429326P.  
PR 27-NOV-2002; 2002US-0429351P.  
PR 04-DEC-2002; 2002US-0430645P.  
PR 04-DEC-2002; 2002US-0430651P.  
PR 04-DEC-2002; 2002US-0430657P.  
PR 04-DEC-2002; 2002US-0430663P.  
PR 04-DEC-2002; 2002US-0430686P.  
PR 05-DEC-2002; 2002US-0430684P.  
PR 05-DEC-2002; 2002US-0430937P.  
PR 05-DEC-2002; 2002US-0430965P.  
PR 12-DEC-2002; 2002US-0431458P.  
PR 12-DEC-2002; 2002US-0433251P.  
PR 13-DEC-2002; 2002US-0433300P.  
PR 13-DEC-2002; 2002US-0433316P.  
PR 13-DEC-2002; 2002US-0433318P.  
PR 23-DEC-2002; 2002US-0436236P.  
PR 03-JAN-2003; 2003US-0437914P.  
PR 17-JAN-2003; 2003US-0440820P.  
PR 17-JAN-2003; 2003US-0440821P.  
PR 18-APR-2003; 2003US-0463700P.  
PR 18-APR-2003; 2003US-0463708P.  
PR 18-APR-2003; 2003US-0463716P.  
PR 18-APR-2003; 2003US-0463732P.  
PR 02-MAY-2003; 2003US-0467199P.  
PR 02-MAY-2003; 2003US-0467201P.  
PR 02-MAY-2003; 2003US-0467203P.  
PR 02-MAY-2003; 2003US-0467230P.  
PR 19-MAY-2003; 2003US-0471306P.  
PR 19-MAY-2003; 2003US-0471336P.  
PR 22-MAY-2003; 2003US-0472420P.  
PR 22-MAY-2003; 2003US-0472430P.  
PR 09-JUN-2003; 2003US-0476609P.  
PR 09-JUN-2003; 2003US-0476621P.  
PR 09-JUN-2003; 2003US-0476632P.  
PR 09-JUN-2003; 2003US-0476641P.  
PR 08-JUL-2003; 2003US-0485217P.  
PR 08-JUL-2003; 2003US-0485218P.  
PR 08-JUL-2003; 2003US-0485233P.  
PR 08-JUL-2003; 2003US-0485224P.  
PR 08-JUL-2003; 2003US-0485325P.  
PR 08-JUL-2003; 2003US-0485359P.  
PR 14-JUL-2003; 2003US-0486446P.  
PR 14-JUL-2003; 2003US-0486480P.  
PR 15-JUL-2003; 2003US-0486891P.  
PR 15-JUL-2003; 2003US-0486960P.  
PR 08-AUG-2003; 2003US-0493341P.  
PR 08-AUG-2003; 2003US-0493370P.  
PR 08-AUG-2003; 2003US-0493573P.  
PR 08-AUG-2003; 2003US-0493577P.

(FIVE-) FIVE PRIME THERAPEUTICS INC.

PA Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;  
PI Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;  
PI Wong JGP, Wu G, Zhang H, Zeng C;

WPI; 2004-365511/34.

DR N-P5DB; ADN98657.

XX New nucleic acid molecules, useful in preparing a composition for  
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral  
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or  
PT ulcerative colitis.

PS Claim 14; SEQ ID NO 1041; 532pp; English.

XX The invention relates to a nucleic acid molecule comprising a  
CC polynucleotide sequence or its complement that encodes a polypeptide. The  
CC nucleic acid is useful in preparing a composition for treating or  
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,

CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic  
CC heart disease or ulcerative colitis. This sequence corresponds to a  
CC protein of the invention.  
XX  
XX  
SQ Sequence 180 AA;

Query Match 93.9%; Score 31; DB 8; Length 180;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7  
Db 138 AEAESMY 144  
|||||:|

RESULT 12  
ABM80639  
ID ABM80639 standard; protein; 198 AA.  
XX  
AC ABM80639;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Tumour-associated antigenic target (TAT) polypeptide PRO81266, SEQ:1661.  
XX  
DE Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN W02004030615-A2.  
XX  
PD 15-APR-2004.  
XX  
PF 29-SEP-2003; 2003WO-US028547.  
XX  
PR 02-OCT-2002; 2002US-0414971P.  
XX  
PR (GETH ) GENENTECH INC.  
XX  
PI Wu TD, Zhang Z, Zhou Y;  
XX  
XX WPI; 2004-347921/32.  
DR N-PSDB; ACN38265.  
XX  
XX New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX  
PS Claim 12; SEQ ID NO 1661; 7273pp; English.  
XX

CC The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central

CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT polypeptide of the invention  
XX  
XX  
SQ Sequence 198 AA;

Query Match 93.9%; Score 31; DB 8; Length 198;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7  
Db 98 AEAESMY 104  
|||||:|

RESULT 13  
ADX73916  
ID ADX73916 standard; protein; 207 AA.  
XX  
AC ADX73916;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polypeptide seqid 43282.  
XX  
XX plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content.  
XX  
OS Unidentified.  
XX  
PN US2004034888-A1.  
XX  
PD 19-FEB-2004.  
XX  
PF 28-APR-2003; 2003US-00425114.  
XX  
PR 06-MAY-1999; 99US-00304517.  
PR 05-NOV-2001; 2001US-00985678.  
XX  
XX (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABAA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX  
XX WPI; 2004-180133/17.  
XX

CC New recombinant DNA construct, useful for improving plant tolerance to  
CC cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
CC pests, for conferring increased resistance to plant disease, or for  
CC improving yield.  
XX  
PS Claim 1; SEQ ID NO 43282; 15pp; English.  
XX

CC The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring

CC increased resistance to plant disease, for producing galactomannan.  
 CC lignin or plant growth regulators, for increasing the rate of homologous  
 CC recombination in plants, for improving yield by modification of  
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC or by providing improved plant growth and development under at least one  
 CC stress condition or for modifying seed oil or protein yield and/or  
 CC content. This is the amino acid sequence of a plant full length insert  
 CC polypeptide that can be used in the recombinant DNA construct of the  
 CC invention.

XX SQ Sequence 207 AA;  
 Query Match 93.9%; Score 31; DB 8; Length 207;  
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 APAESLY 7  
 Db 136 AEASMY 142  
 |||||:  
 |||||:

RESULT 14  
 ID ADX76784 standard; protein; 237 AA.  
 XX AC ADX76784;  
 XX 23-MAR-2006 (revised)  
 DT 21-APR-2005 (first entry)  
 XX Plant full length insert polypeptide seqid 46150.  
 XX plant protectant; plant growth regulant; gene therapy; plant;  
 KW recombinant DNA construct; physical array; plant breeding marker;  
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
 KW growth rate; cell cycle pathway; disease resistance;  
 KW galactomannan production; lignin production; plant growth regulator;  
 KW yield; plant growth; plant development; seed oil; protein yield;  
 KW protein content.

XX Zea mays.  
 OS US2004034888-A1.  
 PN 19-FEB-2004.  
 PD 28-APR-2003; 2003US-00425114.  
 PF 06-MAY-1999; 99US-00304517.  
 PR 05-NOV-2001; 2001US-00985678.  
 XX (LIU//) LIU J.  
 PA (ZHOU//) ZHOU Y.  
 PA (KOVA//) KOVALIC D K.  
 PA (SCRE//) SCREEN S E.  
 PA (TABA//) TABASKA J E.  
 PA (CAOY//) CAO Y.  
 PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
 XX WPI; 2004-180133/17.  
 DR New recombinant DNA construct, useful for improving plant tolerance to  
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
 PT pests, for conferring increased resistance to plant disease, or for  
 PT improving yield.  
 XX Claim 1; SEQ ID NO 46150; 15pp; English.  
 PS The invention describes a recombinant DNA construct comprising a  
 CC polynucleotide consisting of a sequence encoding an amino acid sequence  
 CC available in electronic form from the US patent office at

CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
 CC of the invention are also useful in physical arrays of molecules and as  
 CC plant breeding markers. The recombinant DNA construct is useful for  
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
 CC plant cells by modification of the cell cycle pathway, for conferring  
 CC increased resistance to plant disease, for producing galactomannan,  
 CC lignin or plant growth regulators, for increasing the rate of homologous  
 CC recombination in plants, for improving yield by modification of  
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC or by providing improved plant growth and development under at least one  
 CC stress condition or for modifying seed oil or protein yield and/or  
 CC content. This is the amino acid sequence of a plant full length insert  
 CC polypeptide that can be used in the recombinant DNA construct of the  
 CC invention.

CC Revised record issued on 23-MAR-2006 : Corrected organism line  
 XX SQ Sequence 237 AA;  
 Query Match 93.9%; Score 31; DB 8; Length 237;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
 Db 194 AEASMY 200  
 |||||:  
 |||||:

RESULT 15  
 ID ABG26376 standard; protein; 273 AA.  
 XX ABG26376  
 AC ABG26376;  
 XX 18-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #26367.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US008631.  
 PF 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS90563.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 56735; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 273 AA;

Query Match 93.9%; Score 31; DB 4; Length 273;  
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7  
 |||||  
 Db 264 AEAESMY 270

Search completed: July 5, 2006, 19:08:52  
 Job time : 65.6042 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:52:16 ; Search time 37.0417 Seconds  
(without alignments)  
87.537 Million cell updates/sec

Title: US-10-774-928A-5

Perfect score: 33

Sequence: 1 AEASLY 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	7	5	US-10-774-928-5
2	33	100.0	10	5	US-10-774-928-7
3	33	100.0	12	4	US-10-444-575-16
4	33	100.0	43	4	US-10-437-963-194037
5	33	100.0	202	4	US-10-437-963-162621
6	33	100.0	297	5	US-10-739-930-10272
7	33	100.0	644	5	US-10-723-860-366
8	33	100.0	5303	6	US-11-097-143-30390
9	31	93.9	125	5	US-10-450-763-51106
10	31	93.9	207	4	US-10-425-114-43282
11	31	93.9	220	4	US-10-437-963-135290
12	31	93.9	220	4	US-10-425-115-365535
13	31	93.9	237	4	US-10-425-114-46150
14	31	93.9	273	5	US-10-450-763-56735
15	31	93.9	275	5	US-10-450-763-45641
16	31	93.9	279	6	US-11-013-684-21
17	31	93.9	398	5	US-10-883-020-5
18	31	93.9	422	3	US-09-779-307-18
19	31	93.9	429	5	US-10-450-763-49905
20	31	93.9	429	5	US-10-450-763-49906
21	31	93.9	431	4	US-10-408-765A-2629
22	31	93.9	476	3	US-09-779-307-16
23	31	93.9	477	3	US-09-779-307-4
24	31	93.9	481	4	US-10-080-334-54
25	31	93.9	481	6	US-11-156-300-24
26	31	93.9	482	3	US-09-779-307-17
27	31	93.9	482	4	US-10-080-334-187

28 31 93.9 482 5 US-10-883-020-1 Sequence 1, Appli  
29 31 93.9 483 3 US-09-919-497-79 Sequence 79, Appli  
30 31 93.9 483 3 US-09-974-298-41 Sequence 41, Appli  
31 31 93.9 483 3 US-09-981-353-158 Sequence 158, App  
32 31 93.9 483 4 US-10-177-293-258 Sequence 258, App  
33 31 93.9 483 4 US-10-080-334-183 Sequence 183, App  
34 31 93.9 483 4 US-10-080-334-184 Sequence 184, App  
35 31 93.9 483 4 US-10-080-334-185 Sequence 185, App  
36 31 93.9 483 4 US-10-080-334-186 Sequence 186, App  
37 31 93.9 483 4 US-10-734-564-116 Sequence 116, App  
38 31 93.9 483 6 US-11-037-713-9 Sequence 9, Appli  
39 31 93.9 500 5 US-10-450-763-39773 Sequence 39773, A  
40 31 93.9 513 5 US-10-450-763-38492 Sequence 38492, A  
41 31 93.9 521 3 US-09-976-782-8 Sequence 8, Appli  
42 31 93.9 929 5 US-10-450-763-39772 Sequence 39772, A  
43 31 93.9 1301 5 US-10-450-763-45996 Sequence 45996, A  
44 30 90.9 55 3 US-09-864-761-41289 Sequence 41289, A  
45 30 90.9 96 5 US-10-501-282-2186 Sequence 2186, Ap

#### ALIGNMENTS

##### RESULT 1

US-10-774-928-5  
; Sequence 5, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-5

Query Match 100.0%; Score 33; DB 5; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.9e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
|||  
Db 1 AEASLY 7

##### RESULT 2

US-10-774-928-7  
; Sequence 7, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-7

Query Match 100.0%; Score 33; DB 5; Length 10;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 1 AEASLY 7

## RESULT 3

US-10-444-575-16  
; Sequence 16, Application US/10444575  
; Publication No. US20030232374A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Connecticut Health Center  
; APPLICANT: Kuchel, George A  
; APPLICANT: Zhu, Qing  
; TITLE OF INVENTION: Compositions and Methods Relating to Detrusor Estrogen-Regulated  
; FILE OF INVENTION: Protein (DERP)  
; FILE REFERENCE: UCT-0035  
; CURRENT APPLICATION NUMBER: US/10/444,575  
; PRIOR FILING DATE: 2003-05-22  
; PRIOR APPLICATION NUMBER: US 60/382,830  
; PRIOR FILING DATE: 2002-05-23  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 16  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-10-444-575-16

Query Match 100.0%; Score 33; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 3 AEASLY 9

## RESULT 4

US-10-437-963-194037  
; Sequence 194037, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 194037  
; LENGTH: 43  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_90118C.1.p  
US-10-437-963-194037

Query Match 100.0%; Score 33; DB 4; Length 43;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 31 AEASLY 37

## RESULT 5

US-10-437-963-162621  
; Sequence 162621, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 162621  
; LENGTH: 202  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_61696C.1.p  
US-10-437-963-162621

Query Match 100.0%; Score 33; DB 4; Length 202;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 131 AEASLY 137

## RESULT 6

US-10-739-930-10272  
; Sequence 10272, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 10272  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-C3304\_1.p  
US-10-739-930-10272

Query Match 100.0%; Score 33; DB 5; Length 297;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 226 AEASLY 232

## RESULT 7

US-10-723-860-366  
; Sequence 366, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsburg, Wendy M.

; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; PRIOR FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8993  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 366  
; LENGTH: 644  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-723-860-366

Query Match 100.0%; Score 33; DB 5; Length 644;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
Db 367 AEASLY 373

## RESULT 8

US-11-097-143-30390  
; Sequence 30390, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.

; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; FILE REFERENCE: DROSOPHILA GENES.

; CURRENT APPLICATION NUMBER: US/11/097,143  
; PRIOR FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30390  
; LENGTH: 5303  
; TYPE: PRT  
; ORGANISM: DROSOPHILA  
US-11-097-143-30390

Query Match 100.0%; Score 33; DB 6; Length 5303;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
Db 3378 AEASLY 3384

## RESULT 9

US-10-450-763-51106

; Sequence 51106, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 51106  
; LENGTH: 125  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: DOMAIN  
; LOCATION: (94)..(125)  
; OTHER INFORMATION: Intermediate filaments proteins domain identified by EMATRIX,  
; OTHER INFORMATION: accession number BL00226C, p-value=9.400e-19, raw score of 13.23  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (65)..(125)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
; OTHER INFORMATION: accession name filament, E-value=1.8e-21, Pfam score of 84.5  
US-10-450-763-51106

Query Match 93.9%; Score 31; DB 5; Length 125;  
Best Local Similarity 85.7%; Pred. No. 81;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
Db 115 AEASLY 121

## RESULT 10

US-10-425-114-43282  
; Sequence 43282, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 43282  
; LENGTH: 207  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700077413\_FLI.pep  
US-10-425-114-43282

Query Match 93.9%; Score 31; DB 4; Length 207;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
Db 136 AEASLY 142

```
RESULT 11
US-10-437-963-135290
; Sequence 135290, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 135290
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_36980C.1.pep
US-10-437-963-135290

Query Match          93.9%; Score 31; DB 4; Length 220;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
DB 149 AEAESMY 155

RESULT 12
US-10-425-115-365535
; Sequence 365535, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 365535
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_96532C.1.pep
US-10-425-115-365535

Query Match          93.9%; Score 31; DB 4; Length 220;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
DB 149 AEAESMY 155

RESULT 13
US-10-425-114-46150
; Sequence 46150, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46150
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701161793_FLI.pep
US-10-425-114-46150

Query Match          93.9%; Score 31; DB 4; Length 237;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
DB 194 AEAESMY 200

RESULT 14
US-10-450-763-56735
; Sequence 56735, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR FILING DATE: 2000-03-31
; PRIOR FILING DATE: 2000-03-31
; PRIOR FILING DATE: 09/540,217
; PRIOR FILING DATE: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 56735
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (177)..(225)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226B, p-values=3.348e-35, raw score of 23.86
US-10-450-763-56735

Query Match          93.9%; Score 31; DB 5; Length 273;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
DB 264 AEAESMY 270
```

RESULT 15  
US-10-450-763-45641  
; Sequence 45641, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 45641  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (179)..(227)  
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,  
; OTHER INFORMATION: accession number BL002268, p-value=3.348e-35, raw score of 23.86  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (87)..(273)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
; OTHER INFORMATION: accession name filament, E-value=3.2e-89, Pfam score of 309.8  
US-10-450-763-45641

Query Match 93.9%; Score 31; DB 5; Length 275;  
Best Local Similarity 85.7%; Pred. NO. 1.9e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7  
|||||:  
Db 266 AEAESNY 272

Search completed: July 5, 2006, 20:57:19  
Job time : 37.0417 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:53:13 ; Search time 3.20833 Seconds  
(without alignments)  
58.543 Million cell updates/sec

Title: US-10-774-928A-5

Perfect score: 33

Sequence: 1 AEAESLY 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New.\*

1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	221	6	US-10-449-902-30681 Sequence 30681, A
2	31	93.9	220	6	US-10-449-902-30877 Sequence 30877, A
3	31	93.9	220	6	US-10-449-902-33167 Sequence 33167, A
4	31	93.9	483	7	US-11-105-233-180 Sequence 180, App
5	30	90.9	282	6	US-10-449-902-54246 Sequence 54246, A
6	30	90.9	474	6	US-10-449-902-31122 Sequence 31122, A
7	30	90.9	520	7	US-11-293-697-3712 Sequence 3712, App
8	30	90.9	527	7	US-10-449-902-54681 Sequence 54681, A
9	30	90.9	637	7	US-11-293-697-2455 Sequence 2455, App
10	30	90.9	754	6	US-10-449-902-41136 Sequence 41136, A
11	29	87.9	514	6	US-10-449-902-52901 Sequence 52901, A
12	29	87.9	514	6	US-10-449-902-53911 Sequence 53911, A
13	28	84.8	95	6	US-10-953-349-53135 Sequence 53135, A
14	28	84.8	217	6	US-10-449-902-44464 Sequence 44464, A
15	27	81.8	214	6	US-10-953-349-2848 Sequence 2848, App
16	27	81.8	346	6	US-10-449-902-40486 Sequence 40486, A
17	27	81.8	428	6	US-10-935-291B-4 Sequence 4, Appli
18	27	81.8	530	7	US-11-174-307B-4746 Sequence 4746, App
19	26	78.8	155	7	US-11-293-697-3730 Sequence 3730, App
20	26	78.8	223	6	US-10-449-902-43747 Sequence 43747, A
21	26	78.8	258	6	US-10-953-349-3597 Sequence 3597, App
22	26	78.8	259	6	US-10-953-349-3596 Sequence 3596, App
23	26	78.8	264	6	US-10-953-349-3595 Sequence 3595, App
24	26	78.8	296	6	US-10-471-571A-2010 Sequence 2010, App
25	26	78.8	348	6	US-10-449-902-35774 Sequence 35774, A

## RESULT 1

US-10-449-902-30681  
; Sequence 30681, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:

; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 30681  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-30681

Query Match 100.0%; Score 33; DB 6; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AEAESLY 7  
DB 150 AEAESLY 156

## RESULT 2

US-10-449-902-30877  
; Sequence 30877, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:

; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269

Sequence 38425, A  
Sequence 22, Appl  
Sequence 11130, A  
Sequence 11129, A  
Sequence 41447, A  
Sequence 11128, A  
Sequence 11389, A  
Sequence 11386, A  
Sequence 4926, A  
Sequence 4942, A  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 25461, A  
Sequence 25460, A  
Sequence 31397, A  
Sequence 33444, A  
Sequence 3848, Ap  
Sequence 33443, A  
Sequence 56102, A  
Sequence 33442, A

```
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30877
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-30877

Query Match      93.9%; Score 31; DB 6; Length 220;
Best Local Similarity 85.7%; Pred. No. 4.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLY 7
DB      149 AEAESMY 155

RESULT 3
US-10-449-902-33167
; Sequence 33167, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33167
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-33167

Query Match      93.9%; Score 31; DB 6; Length 220;
Best Local Similarity 85.7%; Pred. No. 4.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLY 7
DB      149 AEAESMY 155

RESULT 4
US-11-105-233-180
; Sequence 180, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; TITLE OF INVENTION: Tumors
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-180

Query Match      93.9%; Score 31; DB 6; Length 220;
Best Local Similarity 85.7%; Pred. No. 4.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLY 7
DB      149 AEAESMY 155

RESULT 5
US-10-449-902-54246
; Sequence 54246, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54246
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-54246

Query Match      90.9%; Score 30; DB 6; Length 282;
Best Local Similarity 85.7%; Pred. No. 9.7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLY 7
DB      26 AESESLY 32

RESULT 6
US-10-449-902-31122
; Sequence 31122, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31122
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-31122

Query Match      90.9%; Score 30; DB 6; Length 474;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLY 7
```



```
Db          314 AEAESV 320
|||||:-|
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2455
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-2455

Query Match          90.9%; Score 30; DB 7; Length 637;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 AEAESLY 7
|||||:-|
Db          414 AEAETLY 420

RESULT 10
US-10-449-902-41136
; Sequence 41136, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41136
; LENGTH: 754
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-41136

Query Match          90.9%; Score 30; DB 6; Length 754;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 AEAESLY 7
|||||:-|
Db          464 AEAQSLY 470

RESULT 11
US-10-449-902-52901
; Sequence 52901, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870

Query Match          90.9%; Score 30; DB 6; Length 527;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 AEAESLY 7
|||||:-|
Db          367 AEAESVY 373

RESULT 9
US-11-293-697-2455
; Sequence 2455, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE

Query Match          90.9%; Score 30; DB 6; Length 520;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 AEAESLY 7
|||||:-|
Db          298 AEAELY 304

RESULT 8
US-10-449-902-54681
; Sequence 54681, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54681
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-54681

Query Match          90.9%; Score 30; DB 6; Length 527;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY          1 AEAESLY 7
|||||:-|
Db          367 AEAESVY 373

RESULT 9
US-11-293-697-2455
; Sequence 2455, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
```

; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 52901  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-52901

Query Match 87.9%; Score 29; DB 6; Length 514;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
| | | | |  
Db 335 AEAESFY 341

RESULT 12  
US-10-449-902-53911  
; Sequence 53911, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 53911  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-53911

Query Match 87.9%; Score 29; DB 6; Length 514;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
| | | | |  
Db 335 AEAESFY 341

RESULT 13  
US-10-953-349-35135  
; Sequence 35135, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 35135  
; LENGTH: 95  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-35135

Query Match 84.8%; Score 28; DB 6; Length 95;  
Best Local Similarity 85.7%; Pred. No. 8.7;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AEAESLY 7  
| | | | |  
Db 24 AEESLY 30

RESULT 14  
US-10-449-902-44464  
; Sequence 44464, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 44464  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-44464

Query Match 84.8%; Score 28; DB 6; Length 217;  
Best Local Similarity 85.7%; Pred. No. 22;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
| | | | |  
Db 146 AEESLY 152

RESULT 15  
US-10-953-349-2848  
; Sequence 2848, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2848  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-2848

Query Match 81.8%; Score 27; DB 6; Length 214;  
Best Local Similarity 85.7%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
| | | | |  
Db 145 AEESLY 151

Search completed: July 5, 2006, 20:57:47  
Job time : 3.20833 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using 'sw model

Run on: July 5, 2006, 19:24:38 ; Search time 12.3958 Seconds  
(without alignments)  
49.429 Million cell updates/sec

Title: US-10-774-928A-5

Perfect score: 33

Sequence: 1 AEASLY 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5 COMB pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6 COMB pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7 COMB pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H COMB pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/PCTUS COMB pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE COMB pep.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles1 pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	7	2	US-09-809-003A-5
2	33	100.0	10	2	US-09-809-003A-7
3	33	100.0	637	2	US-09-949-016-8152
4	33	100.0	643	2	US-09-538-092-844
5	31	93.9	443	2	US-09-949-016-10582
6	31	93.9	482	2	US-09-538-092-858
7	31	93.9	483	2	US-09-919-497-79
8	31	93.9	502	2	US-09-949-016-11033
9	30	90.9	243	1	US-08-597-236-3
10	30	90.9	243	1	US-08-746-682A-3
11	30	90.9	398	2	US-10-104-047-3388
12	30	90.9	608	2	US-10-094-749-2446
13	30	90.9	619	2	US-10-094-749-1964
14	30	90.9	619	2	US-10-094-749-2525
15	29	87.9	286	2	US-09-252-991A-22234
16	29	87.9	542	2	US-09-302-540-15640
17	29	87.9	735	2	US-09-502-540-13533
18	29	87.9	925	2	US-09-328-352-5244
19	28	84.8	106	2	US-09-248-796A-26131
20	28	84.8	213	1	US-08-531-525-11
21	28	84.8	213	1	US-08-718-270A-11
22	28	84.8	213	2	US-09-134-000C-3952
23	28	84.8	887	2	US-09-540-236-2911
24	28	84.8	1890	2	US-09-502-540-15442
25	27	81.8	107	2	US-09-543-681A-6074
26	27	81.8	171	2	US-09-248-796A-17720

27	81.8	193	2	US-09-543-681A-4479	Sequence 4479, Ap
28	81.8	235	2	US-09-620-405B-487	Sequence 487, App
29	81.8	235	2	US-09-604-287A-487	Sequence 487, App
30	81.8	235	2	US-09-834-759-487	Sequence 487, App
31	81.8	235	2	US-10-076-622-487	Sequence 487, App
32	81.8	235	2	US-10-124-805-487	Sequence 487, App
33	81.8	274	2	US-09-949-016-9095	Sequence 9095, Ap
34	81.8	310	2	US-09-540-236-2984	Sequence 2984, Ap
35	81.8	323	2	US-09-543-681A-6957	Sequence 6957, Ap
36	81.8	384	2	US-09-902-540-10867	Sequence 10867, A
37	81.8	399	1	US-08-096-623A-20	Sequence 20, Appli
38	81.8	422	5	PCT-US91-01360-2	Sequence 4, Appli
39	81.8	428	2	US-10-179-784-4	Sequence 6203, Ap
40	81.8	493	2	US-09-949-016-6203	Sequence 478, App
41	81.8	505	2	US-09-620-405B-478	Sequence 485, App
42	81.8	505	2	US-09-604-287A-478	Sequence 478, App
43	81.8	505	2	US-09-604-287A-485	Sequence 485, App
44	81.8	505	2	US-09-834-759-478	Sequence 478, App
45	81.8	505	2	US-09-834-759-478	Sequence 478, App

#### ALIGNMENTS

#### RESULT 1

US-09-809-003A-5  
; Sequence 5, Application US/09809003A  
; Patent No. 6673351

; GENERAL INFORMATION:

; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission

; TITLE OF INVENTION: Psoriasis

; FILE REFERENCE: 302178

; CURRENT APPLICATION NUMBER: US/09/809,003A

; CURRENT FILING DATE: 2001-03-16

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 5

; LENGTH: 7

; TYPE: PRT

; ORGANISM: Leishmania

US-09-809-003A-5

Query Match 100.0%; Score 33; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7

Db 1 AEASLY 7

#### RESULT 2

US-09-809-003A-7  
; Sequence 7, Application US/09809003A  
; Patent No. 6673351

; GENERAL INFORMATION:

; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission

; TITLE OF INVENTION: Psoriasis

; FILE REFERENCE: 302178

; CURRENT APPLICATION NUMBER: US/09/809,003A

; CURRENT FILING DATE: 2001-03-16

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 7

; LENGTH: 10

; TYPE: PRT

; ORGANISM: Leishmania

US-09-809-003A-7

Query Match 100.0%; Score 33; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.42;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 1 AEASLY 7

RESULT 3  
US-09-949-016-8152  
; Sequence 8152, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8152  
; LENGTH: 637  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-8152

Query Match 100.0%; Score 33; DB 2; Length 637;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 367 AEASLY 373

RESULT 4  
US-09-538-092-844  
; Sequence 844, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurPatSeqFormatter Version 0.9  
; SEQ ID NO 844  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P04264  
US-09-538-092-844

Query Match 100.0%; Score 33; DB 2; Length 643;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7

Db 366 AEASLY 372  
| | | | |

RESULT 5  
US-09-949-016-10582  
; Sequence 10582, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10582  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10582

Query Match 93.9%; Score 31; DB 2; Length 443;  
Best Local Similarity 85.7%; Pred. No. 72;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 236 AEASLY 242

RESULT 6  
US-09-538-092-858  
; Sequence 858, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurPatSeqFormatter Version 0.9  
; SEQ ID NO 858  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P05787  
US-09-538-092-858

Query Match 93.9%; Score 31; DB 2; Length 482;  
Best Local Similarity 85.7%; Pred. No. 79;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLY 7  
| | | | |  
Db 275 AEASLY 281

```
RESULT 7
US-09-919-497-79
; Sequence 79, Application US/09919497
; Patent No. 6773883
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 79
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-79

Query Match          93.9%; Score 31; DB 2; Length 483;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
Db 276 AEAESMY 282

RESULT 8
US-09-949-016-11033
; Sequence 11033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11033
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11033

Query Match          93.9%; Score 31; DB 2; Length 502;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
Db 295 AEAESMY 301

RESULT 9
US-08-597-236-3
; Sequence 3, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
```

```
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-597-236-3

Query Match          90.9%; Score 30; DB 1; Length 243;
Best Local Similarity 85.7%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7
Db 64 AEAELY 70

RESULT 10
US-08-746-682A-3
; Sequence 3, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
```

; REGISTRATION NUMBER: 30256  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 243 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-746-682A-3

Query Match 90.9%; Score 30; DB 1; Length 243;  
Best Local Similarity 85.7%; Pred. No. 63;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
||||:|  
Db 64 AEAEALY 70

## RESULT 11

US-10-104-047-3388  
; Sequence 3388, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. 6943241el full length cdna  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25

; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3388  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-104-047-3388

Query Match 90.9%; Score 30; DB 2; Length 398;  
Best Local Similarity 85.7%; Pred. No. 11e-02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
||||:|  
Db 199 AEAEALY 205

## RESULT 12

US-10-094-749-2446  
; Sequence 2446, Application US/10094749  
; Patent No. 6979557  
; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH cdna  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749

; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2446  
; LENGTH: 608  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-2446

Query Match 90.9%; Score 30; DB 2; Length 608;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
||||:|  
Db 412 AEAEALY 418

## RESULT 13

US-10-094-749-1964  
; Sequence 1964, Application US/10094749  
; Patent No. 6979557  
; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cdna  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1964  
; LENGTH: 619  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-094-749-1964

Query Match 90.9%; Score 30; DB 2; Length 619;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
||||:|  
Db 396 AEAEALY 402

## RESULT 14

US-10-094-749-2525  
; Sequence 2525, Application US/10094749

Patent No. 6979557  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKI, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTYUYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2525  
; LENGTH: 619  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-2525

Query Match 90.9%; Score 30; DB 2; Length 619;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
Db 396 AEAETLY 402

RESULT 15  
US-09-252-991A-23234  
; Sequence 23234, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23234  
; LENGTH: 286  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23234

Query Match 87.9%; Score 29; DB 2; Length 286;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEAESLY 7  
Db 396 AEAETLY 402

Db 88 AEAEXLY 94

Search completed: July 5, 2006, 19:26:43  
Job time: 12.3958 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 14.5631 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928A-6

Perfect score: 69

Sequence: 1 NYSPPYNTIDDL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.\*

1: Piri.\*

2: Piri.\*

3: Piri.\*

4: Piri.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	100.0	622	2 I37984	keratin 9, type I,
2	53	76.8	442	2 A95196	sugar ABC transpor
3	53	76.8	442	2 F98062	hypothetical prote
4	50	72.5	359	2 B26135	keratin, 52K type
5	50	72.5	420	2 A37343	keratin 13, type I
6	50	72.5	458	1 KRHU3	keratin 13, type I
7	50	72.5	472	1 KRHUE	keratin 14, type I
8	50	72.5	473	2 JC4313	keratin 16, type I
9	50	72.5	526	1 KRBOVI	keratin, 54K type
10	50	72.5	561	2 A31994	keratin 10, type I
11	50	72.5	593	1 KRHU0	keratin 10, type I
12	47	68.1	368	2 A28825	keratin, type I no
13	46	66.7	569	1 KRMSI1	keratin, 59K type
14	46	66.7	570	2 S07330	keratin, epidermal
15	44	63.8	1167	2 B71924	cag island protein
16	43	62.3	400	1 KRHU9	keratin 19, type I
17	43	62.3	513	1 S43941	protein kinase DUN
18	43	62.3	702	2 C84111	hypothetical prote
19	43	62.3	1247	2 F71616	hypothetical prote
20	43	62.3	1536	2 S59841	4-alpha-glucanotra
21	43	62.3	1577	2 T30858	glucosyltransferas
22	42	60.9	371	2 A64481	hypothetical prote
23	42	60.9	432	2 S30433	keratin 17, type I
24	42	60.9	437	2 A55682	keratin 13, type I
25	42	60.9	473	4 A33652	probable keratin 1
26	41	59.4	352	2 F70249	hypothetical prote
27	41	59.4	399	2 A25470	cytokeratin 19 - b
28	40	58.0	371	2 F90486	dehydrogenase, pro
29	40	58.0	433	2 S01631	keratin, type I, e

#### ALIGNMENTS

##### RESULT 1

I37984

keratin 9, type I, cytoskeletal - human

N;Alternate names: cytokeratin 9; scatter protein 60K chain, placental

C;Species: Homo sapiens (man)

C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text change 09-Jul-2004

C;Accession: I37984; S40307; S77921; S41161; B35494; I37943

R;Reis, A.; Hennies, H.C.; Langbein, L.; Digweed, M.; Mischke, D.; Drechsler, M.; Schroe

Nature Genet. 6, 174-179, 1994

A;Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).

A;Reference number: I37984; MUID:94214498; PMID:7512862

A;Accession: I37984

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-622 <RES>

A;Cross-references: UNIPROT:P35527; UNIPARC:UPI000012DAE1; EMBL:X75015; NID:9453154; PID:

R;Langbein, L.; Heid, H.W.; Moll, I.; Franke, W.W.

Differentiation 55, 57-71, 1993

A;Title: Molecular characterization of the body site-specific human epidermal cytokeratin

A;Reference number: I37943; MUID:94131202; PMID:7507869

A;Accession: S40307

A;Molecule type: mRNA

A;Residues: 1-622 <LAN>

A;Cross-references: UNIPARC:UPI000012DAE1; EMBL:Z29074; NID:9435475; PIDN:CAA82315.1; PI:

A;Accession: S77921

A;Molecule type: protein

A;Residues: 29-53;62-104;167-188;199-233;241-249;295-362;374-430;450-480;579-604 <LAF>

A;Cross-references: UNIPARC:UPI00001774A9; UNIPARC:UPI00001774AB; UNIPARC:UPI00001774AB;

4B0; UNIPARC:UPI00001774B1

R;Langbein, L.

submitted to the EMBL Data Library, December 1993

A;Reference number: S41161

A;Accession: S41161

A;Molecule type: mRNA

A;Residues: 1-11, SR, 13-622 <LAW>

A;Cross-references: UNIPARC:UPI000016ABA2; EMBL:Z29074; NID:9435475; PIDN:CAA82315.1; PI:

R;Rosen, E.M.; Meromsky, L.; Romero, R.; Setter, E.; Goldberg, I.

Biochem. Biophys. Res. Commun. 168, 1082-1088, 1990

A;Title: Human placenta contains an epithelial scatter protein.

A;Reference number: A35494; MUID:90267446; PMID:2140676

A;Accession: B35494

A;Molecule type: protein

A;Residues: 'X', 450-465 <ROS>

A;Cross-references: UNIPARC:UPI00001774B2

C;Genetics:

A;Gene: GDB:KRT9; EPPK

A;Cross-references: GDB:303970; OMIM:144200

A;Map position: 17q12-17q21

A;Introns: 213/3; 241/2; 293/3; 347/3; 389/3; 464/2

A;Note: defects in this gene may cause epidermolytic palmoplantar keratoderma

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil; intermediate filament

L1 protein - human  
conserved hypochet  
probable membrane  
hypothetical prote  
tipC protein - sli  
hypothetical prote  
probable transcrip  
periplasmic chaper  
hypothetical prote  
probable transcrip  
keratin 3, type I,  
shopry-associated  
keratin, 47.6K typ  
keratin, 48K type  
keratin 3, type I,  
complement factor

30 40 58.0 51.0 1 S15620  
31 40 58.0 551 2 G95176  
32 40 58.0 754 2 S52816  
33 40 58.0 1105 2 T47582  
34 40 58.0 3848 2 T17414  
35 39 56.5 146 2 F90141  
36 39 56.5 224 2 T51638  
37 39 56.5 227 2 I41062  
38 39 56.5 254 2 E72554  
39 39 56.5 299 2 T47917  
40 39 56.5 327 2 S04511  
41 39 56.5 398 2 A45633  
42 39 56.5 404 2 JS0073  
43 39 56.5 412 1 KRSHL1  
44 39 56.5 486 1 KRXL  
45 39 56.5 764 1 BBHU

F;1-153/Domain: head #status predicted <HEA>  
F;154-459/Domain: helical rod #status predicted <ROD>  
F;460-622/Domain: tail #status predicted <TAI>

Query Match 100.0%; Score 69; DB 2; Length 622;  
Best Local Similarity 100.0%; Pred. No. 0.00074;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
|||  
Db 199 NYSPPYNTIDDL 210

## RESULT 2

A95196  
sugar ABC transporter, sugar-binding protein SPL683 [imported] - Streptococcus pneumoniae  
C;Species: Streptococcus pneumoniae  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C;Accession: A95196

R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A;Reference number: A95000; MUID:21357209; PMID:11463916

A;Accession: A95196

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-442 <KUR>

A;Cross-references: UNIPROT:Q97PB6; UNIPARC:Q8DN08; UNIPARC:UPI0000051948; GB:AE005672;

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SPL683

Query Match 76.8%; Score 53; DB 2; Length 442;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YSPYNTID 10  
|||  
Db 390 YSPYNTID 398

## RESULT 3

F98062  
hypothetical protein ABC-SBP [imported] - Streptococcus pneumoniae (strain R6)  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C;Accession: F98062

R;Hoising, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; B  
y, P.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: F98062

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-442 <KUR>

A;Cross-references: UNIPROT:Q97PB6; UNIPARC:Q8DN08; UNIPARC:UPI0000051948; GB:AE007317;

C;Genetics:

A;Gene: ABC-SBP

Query Match 76.8%; Score 53; DB 2; Length 442;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YSPYNTID 10  
|||  
Db 390 YSPYNTID 398

## RESULT 4

B26135  
keratin, 52K type I epidermal (clone pKSCC 52) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jul-2004  
C;Accession: B26135

R;Knapp, B.; Reintrop, M.; Schweizer, J.; Winter, H.

J. Biol. Chem. 262, 938-945, 1987

A;Title: Three cDNA sequences of mouse type I keratins. Cellular localization of the mRNA

A;Reference number: A92654; MUID:87109202; PMID:2433272

A;Accession: B26135

A;Molecule type: mRNA

A;Residues: 1-359 <KNA>

A;Cross-references: UNIPROT:Q61781; UNIPARC:UPI0000020FF9; GB:M13806; NID:G198628; PIDN:1

C;Superfamily: cytoskeletal keratin

Query Match 72.5%; Score 50; DB 2; Length 359;  
Best Local Similarity 66.7%; Pred. No. 0.73;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
:||||: |||  
Db 42 DYSPYFKTIEDL 53

## RESULT 5

A37343  
keratin 13, type I, cytoskeletal, short splice form - human  
N;Alternate names: cytokeratin 13  
C;Species: Homo sapiens (man)  
C;Date: 19-Mar-1993 #sequence\_revision 19-Mar-1993 #text\_change 09-Jul-2004  
C;Accession: A37343

R;Kuruc, N.; Leube, R.E.; Moll, I.; Bader, B.L.; Franke, W.W.

Differentiation 42, 111-123, 1989

A;Title: Synthesis of cytokeratin 13, a component characteristic of internal stratified

A;Reference number: A37343; MUID:90228645; PMID:2483837

A;Accession: A37343

A;Molecule type: mRNA

A;Residues: 1-420 <KUR>

A;Cross-references: UNIPROT:PI3646; UNIPARC:UPI0000160DOC; GB:X52426; NID:G30376; PIDN:1

C;Genetics:

A;Gene: GDB:KRT13

A;Cross-references: GDB:120740; OMIM:148065

A;Map position: 17q21-17q22

C;Superfamily: cytoskeletal keratin

C;Keywords: alternative splicing; coiled coil; intermediate filament

F;1-96/Domain: head <HEA>

F;97-412/Domain: helical rod #status predicted <ROD>

F;413-420/Domain: tail <TAI>

Query Match 72.5%; Score 50; DB 2; Length 420;  
Best Local Similarity 66.7%; Pred. No. 0.86;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
:||||: |||  
Db 151 DYSPYKTIIEEL 162

## RESULT 6

KRHU3  
keratin 13, type I, cytoskeletal, long splice form - human  
N;Alternate names: cytokeratin 13  
C;Species: Homo sapiens (man)  
C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C;Accession: S06088; A33216; B37343; A33403

R;Mischke, D.; Wächter, E.; Hochstrasser, K.; Wild, A.G.; Schulz, P.

Nucleic Acids Res. 17, 7984, 1989

A;Title: The N-, but not the C-terminal domains of human keratins 13 and 15 are closely

A;Reference number: S06088; MUID:90016882; PMID:2477803

A;Accession: S06088

A;Molecule type: mRNA

A;Residues: 1-458 <MIS1>  
A;Cross-references: UNIPROT:P13646; UNIPARC:UPI0000016AB96; EMBL:X14640; NID:g34032; PIDN  
A;Accession: A33216  
A;Molecule type: protein  
A;Residues: 291-299 <MIS2>  
A;Cross-references: UNIPARC:UPI00000173D50  
R;Kuruc, N.; Leube, R.E.; Moll, I.; Bader, B.L.; Franke, W.W.  
Differentiation 42, 111-123, 1989  
A;Title: Synthesis of cytokeratin 13, a component characteristic of internal stratified  
A;Reference number: A37343; MUID:90228645; PMID:2483837  
A;Accession: B37343  
A;Molecule type: mRNA  
A;Residues: 1-57, 'G', 59-458 <KUR>  
A;Cross-references: UNIPARC:UPI0000012DAE6; GB:X52426  
R;Schulz, P.; Wächter, E.; Hochstrasser, K.; Wild, A.G.; Mischke, D.  
Biochem. Biophys. Res. Commun. 162, 1522-1527, 1989  
A;Title: Sequence of a human keratin 13 specific cDNA encompassing coil 1B through the 3  
A;Reference number: A33403; MUID:89350978; PMID:2475110  
A;Accession: A33403  
A;Molecule type: mRNA  
A;Residues: 158-458 <SCH>  
A;Cross-references: UNIPARC:UPI00000173D51  
C;Genetics:  
A;Gene: GDB:KRT13  
A;Cross-references: GDB:120740; OMIM:148065  
A;Map position: 17q21-17q22  
C;Superfamily: cytoskeletal keratin  
C;Keywords: alternative splicing; coiled coil; intermediate filament  
F;1-96/Domain: head <HEA>  
F;97-412/Domain: helical rod #status predicted <ROD>  
F;413-458/Domain: tail <TAI>

Query Match 72.5%; Score 50; DB 1; Length 458;  
Best Local Similarity 66.7%; Pred. No. 0.94;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NYSPPYNTIDDL 12  
Db 151 DYSPPYKTIEL 162

RESULT 7  
KRHUE  
keratin 14, type I, cytoskeletal - human  
N;Alternate names: 50K type I keratin; cytokeratin 14; epidermal type I keratin  
C;Species: Homo sapiens (man)  
C;Date: 13-Jun-1983 #sequence revision 08-Feb-1996 #text\_change 31-Dec-2004  
C;Accession: A26763; A90858; A90829; I39329; A02939  
R;Marchuk, D.; McCrohon, S.; Fuchs, E.  
Proc. Natl. Acad. Sci. U.S.A. 82, 1609-1613, 1985  
A;Title: Complete sequence of a gene encoding a human type I keratin: sequences homolog  
A;Reference number: A26763; MUID:85166180; PMID:2580298  
A;Accession: A26763  
A;Molecule type: DNA  
A;Residues: 1-472 <MAR1>  
A;Cross-references: UNIPROT:P02533; UNIPROT:Q13092; UNIPARC:UPI0000016AB91; GB:J00124; NI  
R;Marchuk, D.; McCrohon, S.; Fuchs, E.  
Cell 39, 491-498, 1984  
A;Title: Remarkable conservation of structure among intermediate filament genes.  
A;Reference number: A90858; MUID:85074457; PMID:6210150  
A;Accession: A90858  
A;Molecule type: DNA  
A;Residues: 2-472 <MAR2>  
A;Cross-references: UNIPARC:UPI00000173D4B; GB:J00124; NID:g186704; PIDN:AA859562.1; PID:  
A;Note: Initiator Met not shown  
R;Hanukoglu, I.; Fuchs, E.  
Cell 31, 243-252, 1982  
A;Title: The cDNA sequence of a human epidermal keratin: divergence of sequence but cons  
A;Reference number: A90829; MUID:83129327; PMID:6186381  
A;Accession: A90829  
A;Molecule type: mRNA  
A;Residues: 64-93, 'T', 95-472 <HAN>  
A;Cross-references: UNIPARC:UPI00000173D4C; GB:J00124; NID:g186704

R;Albers, K.; Fuchs, E.  
J. Cell Biol. 105, 791-806, 1987  
A;Title: The expression of mutant epidermal keratin cDNAs transfected in simple epitheli  
A;Reference number: I39329; MUID:87308384; PMID:2442174  
A;Accession: I39329  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 336-348, 'V', 350-467, 'DGSSSSASCQFFGLM' <ALB>  
A;Cross-references: UNIPARC:UPI000007119B; GB:M28646; NID:g177138; PIDN:AAA96689.1; PID:  
C;Comment: The mammalian cytokeratins are a heterogeneous group of intermediate filament  
segments with beta conformation, and amino and carboxyl ends with diverse length and se  
C;Comment: The helical core plays a role in the polymerization of two polypeptide chains  
nal specificity to each chain and is responsible for the end-to-end linkage of about 10  
C;Comment: Each end domain consists of a set of subdomains distributed with bilateral sy  
ggesting functional significance. Because they are easily removed by proteolysis, these  
C;Comment: The rod domain is flanked by V1 and V2 subdomains that are highly variable in  
subdomains are strongly basic and are also variable in sequence.  
C;Comment: The cytoskeletal and microfibrillar keratins are classified into two types, t  
atin IF protein subunit appears to be a heterotetramer of two type I and two type II pro  
C;Genetics:  
A;Gene: GDB:KRT14; EBS4; EBS3  
A;Cross-references: GDB:132145; OMIM:148066  
A;Map position: 17q12-17q21  
A;Introns: 175/3; 203/2; 255/3; 309/3; 351/3; 425/2; 441/1  
A;Note: defects in this gene may result in epidermolysis bullosa simplex, Dowling-Meara  
C;Complex: heterotetramer of two type I and two type II proteins, usually keratin 5 (see  
C;Keywords: coiled coil; heterotetramer; intermediate filament  
F;2-472/Product: keratin, 50K type 1 cytoskeletal #status predicted <MAT>  
F;2-115/Domain: head <HED>  
F;2-115/Region: E1 and V1 subdomains  
F;116-426/Domain: rod <ROD>  
F;116-150/Region: coil 1A  
F;151-161/Region: linker 1  
F;162-262/Region: coil 1B  
F;263-278/Region: linker 12  
F;279-297/Region: coil 2A  
F;298-305/Region: linker 2  
F;306-426/Region: coil 2B  
F;364/Region: stutler  
F;427-472/Domain: tail <END>  
F;427-472/Region: V2 and E2 subdomains

Query Match 72.5%; Score 50; DB 1; Length 472;  
Best Local Similarity 66.7%; Pred. No. 0.97;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NYSPPYNTIDDL 12  
Db 161 DYSPPYKTIEL 172

RESULT 8  
JC4313  
keratin 16, type I, cytoskeletal - human  
N;Alternate names: 46K keratin type I; cytokeratin 16  
C;Species: Homo sapiens (man)  
C;Date: 06-Dec-1995 #sequence revision 08-Feb-1996 #text\_change 09-Jul-2004  
C;Accession: JC4313; A24843; I58129  
R;Paladini, R.D.; Takahashi, K.; Gant, T.M.; Coulombe, P.A.  
Biochem. Biophys. Res. Commun. 215, 517-523, 1995  
A;Title: cDNA cloning and bacterial expression of the human type I keratin 16.  
A;Reference number: JC4313; MUID:96011809; PMID:7487986  
A;Accession: JC4313  
A;Molecule type: mRNA  
A;Residues: 1-473 <PAL>  
A;Cross-references: UNIPROT:P08779; UNIPARC:UPI0000016B4A5; GB:S79867; NID:g1195530; PIDN  
A;Experimental source: epidermal keratinocytes  
R;Raychaudhury, A.; Marchuk, D.; Lindhurst, M.; Fuchs, E.  
Mol. Cell. Biol. 6, 539-548, 1986  
A;Title: Three tightly linked genes encoding human type I keratins: conservation of sequ  
A;Reference number: A24843; MUID:87064338; PMID:2431270  
A;Accession: A24843

A:Molecule type: DNA  
A:Residues: 1,'T',3-25,'A','27-37','A','39-40,43,'ASTY',48-49,'A','51-186,'HAL',190-207,'AR'  
A:Cross-references: UNIPARC:UPI00001774B6; GB:M28439; NID:g186683; PIDN:AAA59460.1; PID:  
A:Note: the translated sequence in GenBank entry HUMKER16A8, release 111.0, (PIDN:AAA59460.1)  
R;McLean, W.H.I.; Rugg, E.L.; Lunny, D.P.; Morley, S.M.; Lane, E.B.; Swenson, O.; Doppel  
Kunkeler, L.; Munro, C.S.  
Nature Genet. 9, 273-278, 1995  
A:Title: Keratin 16 and keratin 17 mutations cause pachyonychia congenita.  
A:Reference number: I58129; MUID:95291318; PMID:7539673  
A:Accession: I58129  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 118-131,'P',133-134 '<MCL>  
A:Cross-references: UNIPARC:UPI000016548F; GB:S78514; NID:g1000376; PIDN:AAB34564.1; PID:  
A:Note: this is a mutant sequence  
C:Comment: This protein is an intermediate filament protein and expressed in epithelial  
ound healing, psoriasis and cancer.  
C:Genetics:  
A:Gene: KRT16  
A:Cross-references: GDB:136207; OMIM:148067  
A:Map position: 17pter-17qter  
A:Note: Defects in this gene may result in Jadassohn-Lewandowsky pachyonychia congenita  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 72.5%; Score 50; DB 2; Length 473;  
Best Local Similarity 66.7%; Pred.No. 0.98;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NYSPPYNTIDDL 12  
:||||: |||:  
Db 163 DYSPPFKTIEDL 174

RESULT 9  
KRBOVI  
keratin, 54K type I cytoskeletal - bovine  
N;Alternate names: 54-kDa type I keratin, cytokeratin VIB  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 09-Jul-2004  
C:Accession: A02941; S07262  
R;Rieger, M.; Jorcano, J.L.; Franke, W.W.  
EMBO J. 4, 2261-2267, 1985  
A:Title: Complete sequence of a bovine type I cytokeratin gene: conserved and variable i  
A:Reference number: A02941; MUID:86081734; PMID:2416562  
A:Accession: A02941  
A:Molecule type: DNA  
A:Residues: 1-526 <RIE>  
A:Cross-references: UNIPROT:P06394; UNIPARC:UPI000012DAE2; GB:X02870; NID:g478; PIDN:CAP  
R;Jorcano, J.L.; Rieger, M.; Franke, J.K.; Schiller, D.L.; Moil, R.; Franke, W.W.  
J. Mol. Biol. 179, 257-281, 1984  
A:Title: Identification of two types of keratin polypeptides within the acidic cytokerat  
A:Reference number: S07262; MUID:85058191; PMID:6209405  
A:Accession: S07262  
A:Molecule type: mRNA  
A:Residues: 281-466,'PAAATAAEVQRRPRXFRQYQ',490-491,494,'PVAVARRRK',504-526 <JOR>  
A:Cross-references: UNIPARC:UPI0000173DS6; EMBL:X01460  
A:Note: this sequence has been revised in reference A02941  
C:Genetics:  
A:Gene: cytokeratin VIB  
A:Introns: 190/3; 218/2; 270/3; 324/3; 366/3; 439/2; 525/2  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; intermediate filament  
F:1-127/Domain: head <HED>  
F:1-127/Region: E1 and V1 subdomains  
F:128-441/Domain: rod <ROD>  
F:128-162/Region: coil 1A  
F:163-176/Region: linker 1  
F:177-277/Region: coil 1B  
F:278-293/Region: linker 12  
F:294-312/Region: coil 2A  
F:313-320/Region: linker 2  
F:321-441/Region: coil 2B

F;379/Region: stutter  
F;442-526/Domain: tail <END>  
F;442-526/Region: V2 and E2 subdomains

Query Match 72.5%; Score 50; DB 1; Length 526;  
Best Local Similarity 75.0%; Pred. No. 1.1;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NYSPTYNTIDDL 12  
Db 176 DYSKYQTIDDL 187  
:  
||| |||  
||| |||

RESULT 10  
A31994  
keratin 10, type I, epidermal - human  
N;Alternate names: keratin 10 V2 variant 117  
C;Species: Homo sapiens (man)  
C;Date: 21-May-1990 #sequence\_revision 21-May-1990 #text\_change 09-Jul-2004  
C;Accession: A31994; A3182  
J;Zhong, X.M.; Idter, W.W.; Steven, A.C.; Roop, D.R.; Steinert, P.M.  
J. Biol. Chem. 263, 15584-15589, 1988  
A;Title: The complete sequence of the human intermediate filament chain keratin 10. Subd  
A;Reference number: A31994; MUID:89008465; PMID:2459124  
A;Accession: A31994  
A:Molecule type: mRNA  
A;Residues: 1-561 <ZHO>  
A;Cross-references: UNIPROT:Q14664; UNIPARC:UPI0000073D6C; GB:J04029; NID:g623408; PIDN:f  
R;Korge, B.P.; Gan, S.Q.; McBride, O.W.; Mischke, D.; Steinert, P.M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 910-914, 1992  
A;Title: Extensive size polymorphism of the human keratin 10 chain resides in the C-termi  
A;Reference number: A38182; MUID:92141228; PMID:1371013  
A;Accession: A38182  
A;Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A;Residues: 445-561 <KOR>  
A;Cross-references: UNIPARC:UPI00001774B3  
A;Note: sequence extracted from NCBI backbone (NCBIP:79433)  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 72.5%; Score 50; DB 2; Length 561;  
Best Local Similarity 75.0%; Pred. No. 1.2;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NYSPTYNTIDDL 12  
Db 188 DYSKYKTIDDL 199  
:  
||| |||  
||| |||

RESULT 11  
KRHUO  
keratin 10, type I, cytoskeletal - human  
N;Alternate names: cytokeratin 10  
C;Species: Homo sapiens (man)  
C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C;Accession: S02158; C38182; B38182; PC1102; S14666; S14669  
R;Rieger, M.; Franke, W.W.  
J. Mol. Biol. 204, 841-856, 1988  
A;Title: Identification of an orthologous mammalian cytokeratin gene. High degree of inte  
A;Reference number: S02158; MUID:89125611; PMID:2464696  
A;Accession: S02158  
A:Molecule type: DNA  
A;Residues: 1-593 <RIE>  
A;Cross-references: UNIPROT:P13645; UNIPARC:UPI000016A483; EMBL:X14487; NID:g28316; PIDN  
A;Experimental source: clone lambda-KH10-5  
R;Korge, B.P.; Gan, S.Q.; McBride, O.W.; Mischke, D.; Steinert, P.M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 910-914, 1992  
A;Title: Extensive size polymorphism of the human keratin 10 chain resides in the C-termi  
A;Reference number: A38182; MUID:92141228; PMID:1371013  
A;Accession: C38182  
A;Status: not compared with conceptual translation  
A:Molecule type: DNA

```

QY      I NISPYINIDDL 12
      :|| || ||: ||
Db      192 DYSKYKTIEDL 203

RESULT 14
S07330  keratin, epidermal - mouse
C;Species: Mus musculus (house mouse)
C;Date: 01-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S07330
R;Steinert, P.M.; Rice, R.H.; Roop, D.R.; Trus, B.L.; Steven, A.C.
Nature 302, 794-800, 1993
A;Title: Complete amino acid sequence of a mouse epidermal keratin subunit and implicati

```

A;Reference number: S07330; MUID:83192464; PMID:6188955  
A;Accession: S07330  
A;Molecule type: mRNA  
A;Residues: 1-570 <STE>  
A;Cross-references: UNIPROT:P02535; UNIPARC:UPI000016CE76; EMBL:V00830; NID:952786; PIDN  
A;Note: the authors translated the codon TTC for residue 25 as Ser, TTC for residue 29 as  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; intermediate filament

Query Match 66.7%; Score 46; DB 2; Length 570;  
Best Local Similarity 66.7%; Pred. No. 5.8;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
Db :||| ||| |||  
192 DYSKYKTIEDL 203

## RESULT 15

B71924  
cag island protein, cytotoxicity associated immunodominant antigen - Helicobacter pylori  
C;Species: Helicobacter pylori  
A;Variety: strain J99  
C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C;Accession: B71924  
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A;Reference number: A71800; MUID:99120557; PMID:9923682  
A;Accession: B71924  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1167 <ARN>  
A;Cross-references: UNIPROT:Q9ZLT1; UNIPARC:UPI0000126DD7; GB:AE001483; GB:AE001439; NID  
A;Experimental source: strain J99  
C;Genetics:  
A;Gene: caga

Query Match 63.8%; Score 44; DB 2; Length 1167;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
Db :||| ||| |||  
948 NHEPIYATIDDL 959

Search completed: July 5, 2006, 19:20:31  
Job time : 15.5631 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 122.796 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-6

Perfect score: 69

Sequence: 1 NYSPPYNTIDDL 12

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	100.0	623	1 K1C9 HUMAN	P35527 homo sapien
2	61	88.4	13	2 Q9UC27 HUMAN	Q9UC27 homo sapien
3	60	87.0	431	2 Q566F9 XENLA	Q566F9 xenopus lae
4	60	87.0	431	2 Q6GLQ7 XENLA	Q6GLQ7 xenopus lae
5	55	79.7	786	2 O18740 CANFA	O18740 canis famil
6	53	76.8	61	2 Q879M7 STRP3	Q879M7 streptococc
7	53	76.8	172	2 Q3DHY3 STRAG	Q3DHY3 streptococc
8	53	76.8	219	2 Q3DI09 STRAG	Q3DI09 streptococc
9	53	76.8	438	2 Q3D3U6 STRAG	Q3D3U6 streptococc
10	53	76.8	438	2 Q3DBH3 STRAG	Q3DBH3 streptococc
11	53	76.8	438	2 Q3DFK5 STRAG	Q3DFK5 streptococc
12	53	76.8	438	2 Q3DL75 STRAG	Q3DL75 streptococc
13	53	76.8	438	2 Q3DSV5 STRAG	Q3DSV5 streptococc
14	53	76.8	438	2 Q3K3Z3 STRA1	Q3K3Z3 streptococc
15	53	76.8	438	2 Q8E2F4 STRA3	Q8E2F4 streptococc
16	53	76.8	438	2 Q8E7W2 STRA3	Q8E7W2 streptococc
17	53	76.8	439	2 Q48VD5 STRPM	Q48VD5 streptococc
18	53	76.8	439	2 Q5XDY4 STRP6	Q5XDY4 streptococc
19	53	76.8	439	2 Q7CN18 STRP8	Q7CN18 streptococc
20	53	76.8	439	2 Q9A1I9 STRP1	Q9A1I9 streptococc
21	53	76.8	442	2 Q8DN08 STRR6	Q8DN08 streptococc
22	53	76.8	442	2 Q97PE6 STRPN	Q97PE6 streptococc
23	51	73.9	618	2 Q8CIS9 RAT	Q8CIS9 rattus norv
24	50	72.5	85	2 Q2TAZ7 HUMAN	Q2TAZ7 homo sapien
25	50	72.5	403	2 Q5NZP1 AZOSE	Q5NZP1 azoarcus sp
26	50	72.5	449	2 Q2MIG8 MOUSE	Q2MIG8 mus musculu
27	50	72.5	452	2 Q6SEK1 MOUSE	Q6SEK1 mus musculu
28	50	72.5	452	2 Q6IFU7 RAT	Q6IFU7 rattus norv
29	50	72.5	452	2 Q6IFX2 MOUSE	Q6IFX2 mus musculu
30	50	72.5	458	1 K1C13 HUMAN	P13646 homo sapien
31	50	72.5	467	1 K1C14_CHICK	Q6P2V1 gallus gall

32	50	72.5	468	1 K1C16 MOUSE	Q9Z2K1 mus musculu
33	50	72.5	469	2 Q3SYP5 MOUSE	Q3SYP5 mus musculu
34	50	72.5	470	2 Q3ZAW8 MOUSE	Q3ZAW8 mus musculu
35	50	72.5	471	1 K1C14 HUMAN	P02533 homo sapien
36	50	72.5	472	1 K1C16 HUMAN	P08779 homo sapien
37	50	72.5	473	2 Q9EQD7 MOUSE	Q9EQD7 mus musculu
38	50	72.5	474	2 Q9EQD6 MOUSE	Q9EQD6 mus musculu
39	50	72.5	484	1 K1C14 MOUSE	Q61781 mus musculu
40	50	72.5	485	1 K1C14 RAT	Q61FV1 rattus norv
41	50	72.5	526	1 K1C10 BOVIN	P06394 bos taurus
42	50	72.5	561	2 Q14664 HUMAN	Q14664 homo sapien
43	50	72.5	584	2 Q8N175 HUMAN	Q8N175 homo sapien
44	50	72.5	593	1 K1C10 HUMAN	P13645 homo sapien
45	49	71.0	127	2 Q5AH45 CANAL	Q5AH45 candida alb

#### ALIGNMENTS

##### RESULT 1

K1C9\_HUMAN  
ID K1C9\_HUMAN STANDARD; PRT; 623 AA.  
AC P35527; O00109; Q14665;  
DT 01-JUN-1994, integrated into UniprotKB/Swiss-Prot.  
DT 08-NOV-2005, sequence version 2.  
DT 07-FEB-2006, entry version 50.  
DE Keratin, type I cytoskeletal 9 (Cytokeratin-9) (CK-9) (Keratin-9)  
DE (K9).  
GN Name=KRT9;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND PARTIAL PROTEIN SEQUENCE.  
RC TISSUE=Foot sole tissue;  
RX MEDLINE=94131202; PubMed=7507869;  
RA Langbein L., Heid H.W., Moll I., Franke W.W.;  
RT "Molecular characterization of the body site-specific human epidermal  
cytochrome 9: cDNA cloning, amino acid sequence, and tissue  
specificity of gene expression.";  
RL Differentiation 55:57-72(1993).  
[2]  
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANTS EPPK LYS-161; GLN-163  
RP AND TRP-163.  
RX MEDLINE=94214498; PubMed=7512862;  
RA Reis A., Hennies H.-C., Langbein L., Digweed M., Mischke D.,  
Dreschler M., Schroeck E., Royer-Pokora B., Franke W.W., Sperling K.,  
Kuester W.;  
RT "Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma  
(EPPK).";  
RL Nat. Genet. 6:174-179(1994).  
[3]  
RP PROTEIN SEQUENCE OF 14-29, AND MASS SPECTROMETRY.  
RC TISSUE=Cervix carcinoma;  
RA Bienvenut W.V.;  
RL Unpublished observations (AUG-2005).  
[4]  
RN NUCLEOTIDE SEQUENCE [MRNA] OF 147-372, AND VARIANT EPPK GLN-163.  
RX MEDLINE=96228052; PubMed=8647270; DOI=10.1016/0014-5793(96)00393-6;  
RA Kobayashi S., Tanaka T., Matsuyoshi N., Inamura S.;  
RT "Keratin 9 point mutation in the pedigree of epidermolytic hereditary  
palmoplantar keratoderma perturbs keratin intermediate filament  
network formation.";  
RL FEBS Lett. 386:149-155(1996).  
[5]  
RN PROTEIN SEQUENCE OF 450-466.  
RX MEDLINE=90267446; PubMed=2140676;  
RA Rosen E.M., Meromsky L., Romero R., Setter E., Goldberg I.;  
RT "Human placenta contains an epithelial scatter protein.";  
RL Biochem. Biophys. Res. Commun. 168:1082-1088(1990).  
[6]

RP VARIANTS EPPK VAL-157 AND PRO-172.  
RX MEDLINE=94274199; PubMed=7516304; DOI=10.1007/BF00201564;  
RA Hennies H.-C., Zehender D., Kunze J., Kuester W., Reis A.;  
RT "Keratin 9 gene mutational heterogeneity in patients with  
RL epidermolytic palmoplantar keratoderma.";  
RN Hum. Genet. 93:649-654(1994).  
RP VARIANTS EPPK SER-161.  
RX MEDLINE=95015968; PubMed=753528; DOI=10.1111/1523-1747.ep12395570;  
RA Bonifas J.M., Matsumura K., Chen M.A., Berth-Jones J.,  
RX Hutchinson P.E., Zloczower M., Fritsch P.O., Epstein E.H. Jr.;  
RA "Mutations of keratin 9 in two families with palmoplantar  
RT epidermolytic hyperkeratosis.";  
RL J. Invest. Dermatol. 103:474-477(1994).  
RN VARIANTS EPPK THR-163 AND SER-168.  
RX MEDLINE=95164983; PubMed=7532199; DOI=10.1111/1523-1747.ep12666018;  
RA Rothnagel J.A., Wojcik S., Liefer K.M., Dominey A.M., Huber M.,  
RA Hohl D., Roop D.R.;  
RX "Mutations in the IA domain of keratin 9 in patients with  
RT epidermolytic palmoplantar keratoderma.";  
RL J. Invest. Dermatol. 104:430-433(1995).  
RN VARIANTS EPPK VAL-160.  
RX MEDLINE=97348961; PubMed=9204965; DOI=10.1111/1523-1747.ep1276751;  
RA Endo H., Hatamochi A., Shinkai H.;  
RX "A novel mutation of a leucine residue in coil 1A of keratin 9 in  
RT epidermolytic palmoplantar keratoderma.";  
RL J. Invest. Dermatol. 109:113-115(1997).  
RN VARIANTS EPPK THR-157, VAL-157 AND GLN-163.  
RX MEDLINE=99072662; PubMed=9856842;  
RA Coveillo S.P., Irvine A.D., McKenna K.E., Munro C.S., Nevin N.C.,  
RA Smith F.J.D., Utito J., McLean W.H.I.;  
RX "Mutations in keratin 9 in kindreds with epidermolytic palmoplantar  
RT keratoderma and epidemiology in Northern Ireland.";  
RL J. Invest. Dermatol. 111:1207-1209(1998).  
CC -I- FUNCTION: May serve an important special function either in the  
CC mature palmar and plantar skin tissue or in the morphogenic  
CC program of the formation of these tissues.  
CC -I- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
CC -I- TISSUE SPECIFICITY: Expressed in the terminally differentiated  
CC epidermis of palms and soles.  
CC -I- DISEASE: Defects in KRT9 are a cause of epidermolytic palmoplantar  
CC keratoderma (EPPK) (MIM:144200); also abbreviated EPPK. EPPK is  
CC an autosomal dominant disease characterized by diffuse thickening  
CC of the epidermis on the entire surface of palms and soles sharply  
CC bordered with erythematous margins.  
CC -I- MISCELLANEOUS: There are two types of cytoskeletal and  
CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
CC 55 and 56-70 kDa, respectively).  
CC -I- SIMILARITY: Belongs to the intermediate filament family.  
CC -I- CAUTION: Was originally (Ref.4) thought to be a 60 kDa chain of  
CC placental scatter protein.  
CC -I- DATABASE: NAME=Human Intermediate Filament Mutation Database;  
CC WWW="http://www.interfil.org".  
CC  
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC -----  
CC ENBL; Z29074; CAA82315.1; -; mRNA.  
CC ENBL; S69510; AAC60619.1; -; mRNA.  
CC ENBL; X75015; CAA52924.1; -; Genomic DNA.

DR EMBL; AB001594; BAA19418.1; -; mRNA.  
DR PIR; I37984; I37984.  
DR HSSP; P08670; IKG7.  
DR Interact; P35527; -.  
DR Ensemble; ENSG00000171403; Homo sapiens.  
DR HGNC; HGNC:6447; KRT9.  
DR MIM; 144200; phenotype.  
DR MIM; 607606; Gene.  
DR GO; GO:0005208; F:structural constituent of cytoskeleton; TAS.  
DR GO; GO:0008544; P:epidermis development; TAS.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin I.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.  
DR Coiled coil; Direct protein sequencing; Disease mutation;  
KW Intermediate filament; Keratin.  
FT CHAIN 1 623 Keratin, type I cytoskeletal 9.  
FT /FTID=PRO\_0000063640.  
FT REGION 1 152 Head.  
FT REGION 153 461 Rod.  
FT REGION 153 188 Coil 1A.  
FT REGION 189 207 Linker 1.  
FT REGION 208 299 Coil 1B.  
FT REGION 300 322 Linker 12.  
FT REGION 323 461 Coil 2.  
FT REGION 462 623 Tail.  
FT COMPIAS 15 26 Poly-Gly.  
FT VARIANT 157 157 M -> T (in EPPK).  
FT VARIANT 157 157 /FTID=VAR\_010499.  
FT VARIANT 160 160 M -> V (in EPPK).  
FT VARIANT 160 160 /FTID=VAR\_010500.  
FT VARIANT 161 161 L -> V (in EPPK).  
FT VARIANT 161 161 /FTID=VAR\_010501.  
FT VARIANT 161 161 N -> K (in EPPK).  
FT VARIANT 161 161 /FTID=VAR\_003822.  
FT VARIANT 161 161 N -> S (in EPPK).  
FT VARIANT 161 161 /FTID=VAR\_010502.  
FT VARIANT 161 161 N -> Y (in EPPK).  
FT VARIANT 163 163 /FTID=VAR\_010503.  
FT VARIANT 163 163 R -> Q (in EPPK).  
FT VARIANT 163 163 /FTID=VAR\_003823.  
FT VARIANT 163 163 R -> W (in EPPK).  
FT VARIANT 168 168 /FTID=VAR\_003824.  
FT VARIANT 172 172 L -> S (in EPPK).  
FT VARIANT 172 172 /FTID=VAR\_003825.  
FT VARIANT 172 172 Q -> P (in EPPK).  
FT CONFLICT 12 13 SR -> T (in Ref. 1; AAC60619 and 2).  
SQ SEQUENCE 623 AA; 62129 MW; DE1D5A462FF96D10 CRC64;  
Query Match 100.0%; Score 69; DB 1; Length 623;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NYSPYNTIDDL 12  
DB 200 NYSPYNTIDDL 211  
RESULT 2  
ID Q9UC27 HUMAN PRELIMINARY; PRT; 13 AA.  
AC Q9UC27;  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 01-MAY-2000, sequence version 1.  
DT 07-FEB-2006, entry version 8.  
DE 94 kDa epidermal cytokeratin-like protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;



RP PROTEIN SEQUENCES.  
RX MEDLINE=95226590; PubMed=7536047;  
RA Boue F., Duquenne C., Lassealle B., Lefevre A., Finaz C.;  
RT "FLB1, a human protein of epididymal origin that is involved in the  
RL sperm-oocyte recognition process.";  
RL Biol. Reprod. 52:267-278(1995).  
CC  
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
SQ SEQUENCE 13 AA; 1610 MW; 000A683B81B7A45B CRC64;  
  
Query Match 88.4%; Score 61; DB 2; Length 13;  
Best Local Similarity 91.7%; Pred. No. 0.0079;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 NYSPPYNTIDDL 12  
Db |||||  
1 NYSPPYNTIDDL 12  
  
RESULT 3  
Q566F9 XENLA PRELIMINARY; PRT; 431 AA.  
AC Q566F9;  
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.  
DT 10-MAY-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE MGC84388 protein.  
GN Name=MGC84388;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodinae; Xenopus; Xenopus.  
OX NCBI\_TaxID=8355;  
[1]  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Brain;  
RC MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT initiative.";  
RL Dev. Dyn. 225:384-391(2002).  
[2]  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Brain;  
RC MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[3]  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Brain;  
RC Klein S., Gerhard D.S.;

RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; BC093567; AAH93567.1; -; mRNA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPEIKERATIN.  
DR PROSITE; PS00226; IF; UNKNOWN 1.  
SQ SEQUENCE 431 AA; 47704 MW; 13EB6E9019731326 CRC64;  
  
Query Match 87.0%; Score 60; DB 2; Length 431;  
Best Local Similarity 83.3%; Pred. No. 0.45;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 NYSPPYNTIDDL 12  
Db |||||  
131 NYSPPYNTIDDL 142  
  
RESULT 4  
Q6GLQ7 XENLA PRELIMINARY; PRT; 431 AA.  
AC Q6GLQ7;  
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 15.  
DE MGC84388 protein.  
GN Name=MGC84388;  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodinae; Xenopus; Xenopus.  
OX NCBI\_TaxID=8355;  
[1]  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Brain;  
RC MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Brain;  
RC MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
RA Richardson P.;  
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
RT initiative.";  
RL Dev. Dyn. 225:384-391(2002).  
[3]  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Brain;

```
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BC074402; AH74402.1; -; mRNA.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; UNKNOWN_1.
KW Intermediate filament.
SQ SEQUENCE 431 AA; 47723 MW; 0032E596D1B243C8 CRC64;

Query Match 87.0%; Score 60; DB 2; Length 431;
Best Local Similarity 83.3%; Pred. No. 0.45;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12
Db 131 NYSPPYSTIDEL 142

RESULT 5
O18740 CANFA PRELIMINARY; PRT; 786 AA.
AC O18740;
DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.
DT 01-JAN-1998, sequence version 1.
DT 07-FEB-2006, entry version 25.
DE Keratin.
GN Name=KRT9;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Lacharme P., Hitté C., Jouquand S., Priat C., Galibert F.;
RT "Identification and analysis of the dog keratin 9 (KRT9) gene.";
RL Anim. Genet. 9:173-178(1998).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AF000949; AAC26971.1; -; Genomic_DNA.
DR HSSP; P08670; 1GK7.
DR HSSP; P08670; 1GK7.
DR Ensembl; ENSCAFG00000015952; Canis familiaris.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR01248; TYPE1KERATIN.
SQ SEQUENCE 786 AA; 76354 MW; 272AB5425DD09535 CRC64;

Query Match 79.7%; Score 55; DB 2; Length 786;
Best Local Similarity 75.0%; Pred. No. 5.3;
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12
Db 184 DYSPPYDTIDEL 195

RESULT 6
Q879M7_STRP3
```

```
ID Q879M7_STRP3 PRELIMINARY; PRT; 61 AA.
AC Q879M7; Q7CFH2;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Putative sugar transporter sugar binding lipoprotein.
GN OrderedLocustNames=SPe0186, SpYM3 0181,
OS Streptococcus pyogenes serotype M3.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Perkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phase-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BA000034; BAC63281.1; -; Genomic_DNA.
DR EMBL; AE014140; AAM78788.1; -; Genomic_DNA.
DR GO; GO:0005351; F:sugar porter activity; IEA.
KW Complete proteome; Lipoprotein; Sugar transport.
SQ SEQUENCE 61 AA; 7153 MW; A82311231A615BB7 CRC64;

Query Match 76.8%; Score 53; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 0.77;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YSPYNTID 10
Db 11 YSPYNTID 19

RESULT 7
Q3DHV3_STRAG PRELIMINARY; PRT; 172 AA.
AC Q3DHV3;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Sugar ABC transporter, sugar-binding protein (Fragment).
GN Name=ABE; ORFNames=SAL_0058;
OS Streptococcus agalactiae 515.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342614;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=515;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Masiarani V., Cieleszczak M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidsen T.M., Mora M., Scarselli M., Margarit y Ros I.,
```

RA	Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA	Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA	Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA	Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA	Smith S., Uterback T.R., White O., Rubens C.E., Grandi G.,
RA	Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA	Fraser C.M.;
RT	"Genome analysis of multiple pathogenic isolates of Streptococcus
RT	agalactiae: implications for the microbial 'pan-genome'.";
RL	Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC	-/- CAUTION: The sequence shown here is derived from an
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
CC	-----
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	EWBL; AAJF01000204; EAO70053.1; -; Genomic_DNA.
DR	NON_TER 172 172
FT	SEQUENCE 172 AA; 19717 MW; AA3ED19930171D02 CRC64;
QY	Query Match 76.8%; Score 53; DB 2; Length 172;
QY	Best Local Similarity 100.0%; Pred. No. 2.3;
QY	Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	2 YSPYYNTID 10
QY	
QY	143 YSPYYNTID 151
QY	-----
QY	RESULT 8
QY	Q3D109_STRAG
QY	ID Q3D109_STRAG PRELIMINARY; PRT; 219 AA.
QY	AC Q3D109;
QY	DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
QY	DT 07-FEB-2006, sequence version 1.
QY	DE Sugar ABC transporter, sugar-binding protein (Fragment).
QY	GN Name=mal; ORFNames=SAL 0059;
QY	OS Streptococcus agalactiae 515.
QY	OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
QY	OC Streptococcus.
QY	OX NCBI_TaxId=342614;
QY	[1]
QY	RC NUCLEOTIDE SEQUENCE.
QY	RC STRAIN=515;
QY	RX PubMed=16173739; DOI=10.1073/pnas.0506758102;
QY	RA Tettelin H., Maignani V., Cieslewicz M.J., Donati C., Medini D.,
QY	RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
QY	RA DeBoy R.T., Davidson T.M., Mora M., Scarcelli M., Margarit y Ros I.,
QY	RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
QY	RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
QY	RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
QY	RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
QY	RA Smith S., Uterback T.R., White O., Rubens C.E., Grandi G.,
QY	RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
QY	RA Fraser C.M.;
QY	"Genome analysis of multiple pathogenic isolates of Streptococcus
QY	agalactiae: implications for the microbial 'pan-genome'.";
QY	Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
QY	-/- CAUTION: The sequence shown here is derived from an
QY	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
QY	preliminary data.
QY	-----
QY	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
QY	Distributed under the Creative Commons Attribution-NoDerivs License
QY	-----
DR	EWBL; AAJF01000187; EAO70079.1; -; Genomic_DNA.
DR	NON_TER 1 1
FT	SEQUENCE 219 AA; 24808 MW; 03260494A3A49847 CRC64;
QY	Query Match 76.8%; Score 53; DB 2; Length 219;
QY	Best Local Similarity 100.0%; Pred. No. 2.9;

RC STRAIN=COH1;  
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;  
RA Tettelin H., Masiugnani V., Cieslewicz M.J., Donati C., Medini D.,  
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,  
RA DeBoy R.T., Daviden T.M., Mora M., Scarselli M., Margarit y Ros I.,  
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,  
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,  
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,  
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,  
RA Smith S., Uterback T.R., White O., Rubens C.E., Grandi G.,  
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,  
RA Fraser C.M.;  
RT "Genome analysis of multiple pathogenic isolates of Streptococcus  
RT agalactiae: implications for the microbial 'pan-genome'.  
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AAJ01000025; EAO75793.1; -; Genomic DNA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
SQ SEQUENCE 438 AA; 48858 MW; FEB4395F0DAF1171 CRC64;  
  
Query Match 76.8%; Score 53; DB 2; Length 438;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 YSPYNTID 10  
Db |||||||  
387 YSPYNTID 395  
  
RESULT 11  
ID Q3DFK5\_STRAG PRELIMINARY; PRT; 438 AA.  
AC Q3DFK5;  
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 4.  
DE Sugar ABC transporter, sugar-binding protein.  
GN ORFNames=SAM 0020;  
OS Streptococcus agalactiae CJB11.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
CC Streptococcus.  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC  
CC  
CC

DR EMBL; AAJ01000011; EAO73713.1; -; Genomic DNA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
SQ SEQUENCE 438 AA; 48858 MW; FEB4395F0DAF1171 CRC64;  
  
Query Match 76.8%; Score 53; DB 2; Length 438;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 YSPYNTID 10  
Db |||||||  
387 YSPYNTID 395  
  
RESULT 12  
ID Q3DL75\_STRAG PRELIMINARY; PRT; 438 AA.  
AC Q3DL75;  
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 4.  
DE Sugar ABC transporter, sugar-binding protein.  
GN ORFNames=SAL\_0039;  
OS Streptococcus agalactiae 515.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
CC Streptococcus.  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC  
CC  
CC

DR EMBL; AAJ01000029; EAO71192.1; -; Genomic DNA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
SQ SEQUENCE 438 AA; 48858 MW; FEB4395F0DAF1171 CRC64;  
  
Query Match 76.8%; Score 53; DB 2; Length 438;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 YSPYNTID 10  
Db |||||||  
387 YSPYNTID 395  
  
RESULT 13  
ID Q3DSY5\_STRAG PRELIMINARY; PRT; 438 AA.  
AC Q3DSY5;  
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 4.  
DE Bacterial extracellular solute-binding protein domain protein.  
CC

```
GN ORNames=SAJ_0091;
OS Streptococcus agalactiae 18RS21.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342613;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=18RS21;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Maignani V., Cieslewicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidson T.M., Mora M., Scarselli M., Margarit y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RT "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ001000037; EA062652.1; -; Genomic DNA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
SQ SEQUENCE 438 AA; 48902 MW; 480FAD6B0A0F30B5 CRC64;

Query Match 76.8%; Score 53; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YSPYNTID 10
Db 387 YSPYNTID 395

RESULT 14
Q3K323 STRA1 PRELIMINARY; PRT; 438 AA.
AC Q3K323;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 4.
DE Carbohydrate uptake 1 (CUT1) family ABC transporter, carbohydrate-
DE binding protein.
GN Streptococcus agalactiae serotype Ia.
OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=355315;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=A909 / ATCC 27591 / Serotype Ia;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Maignani V., Cieslewicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidson T.M., Mora M., Scarselli M., Margarit y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RT "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ001000037; EA062652.1; -; Genomic DNA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
SQ SEQUENCE 438 AA; 48902 MW; 480FAD6B0A0F30B5 CRC64;

Query Match 76.8%; Score 53; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YSPYNTID 10
Db 387 YSPYNTID 395

RESULT 15
Q8E2F4 STRA5 PRELIMINARY; PRT; 438 AA.
AC Q8E2F4;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 15.
DE Sugar ABC transporter, sugar-binding protein.
GN OrderedLocustNames=SAG0034;
OS Streptococcus agalactiae serotype V.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2603 V/R / ATCC BAA-611 / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE014193; AAM98942.1; -; Genomic DNA.
DR HSP; P02928; ILAX.
DR TIGR; SAG0034; -
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
KW Complete proteome.
SQ SEQUENCE 438 AA; 48858 MW; FEB4395F0DAF1171 CRC64;

Query Match 76.8%; Score 53; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YSPYNTID 10
Db 387 YSPYNTID 395
```

```
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000114; ABA45153.1; -; Genomic DNA.
DR TIGR; SAK_0067; -
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
KW Complete proteome.
SQ SEQUENCE 438 AA; 48904 MW; 71B5BCC655254106 CRC64;

Query Match 76.8%; Score 53; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YSPYNTID 10
Db 387 YSPYNTID 395

RESULT 15
Q8E2F4 STRA5 PRELIMINARY; PRT; 438 AA.
AC Q8E2F4;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 15.
DE Sugar ABC transporter, sugar-binding protein.
GN OrderedLocustNames=SAG0034;
OS Streptococcus agalactiae serotype V.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2603 V/R / ATCC BAA-611 / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE014193; AAM98942.1; -; Genomic DNA.
DR HSP; P02928; ILAX.
DR TIGR; SAG0034; -
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
KW Complete proteome.
SQ SEQUENCE 438 AA; 48858 MW; FEB4395F0DAF1171 CRC64;

Query Match 76.8%; Score 53; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YSPYNTID 10
Db 387 YSPYNTID 395
```

Search completed: July 5, 2006, 19:01:34  
Job time : 123.896 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 104.272 Seconds

(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928A-6

Perfect score: 69

Sequence: 1 NYSPYNTIDDL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq.8.\*

1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*  
10: Geneseq2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	69	100.0	12	6 AAO26386	Aao26386 Psoriasis
2	69	100.0	12	9 ADV04411	Adv04411 Leishmani
3	69	100.0	13	4 AAU68343	Aau68343 Human Bre
4	69	100.0	27	10 AEE36578	Aee36578 Human ser
5	69	100.0	622	8 ADS88365	Ads88365 Human pro
6	69	100.0	622	9 AEA15476	Aea15476 Human pol
7	69	100.0	622	9 AEB13340	Aeb13340 Mutant ke
8	65	94.2	600	7 ADF55631	Adf55631 Protein (
9	53	76.8	420	2 AAWS5075	Aaw55075 Streptoco
10	53	76.8	420	5 ABP54569	Abp54569 S. pneumo
11	53	76.8	420	7 ADC45107	Adc45107 S. pneumo
12	53	76.8	438	3 AAY91342	Aay91342 Group B S
13	53	76.8	438	5 ABP25884	Abp25884 Streptoco
14	53	76.8	438	5 ABP29681	Abp29681 Streptoco
15	53	76.8	438	8 ADV89335	Adv89335 Streptoco
16	53	76.8	438	8 ADV82758	Adv82758 Streptoco
17	53	76.8	438	8 ADV80588	Adv80588 Streptoco
18	53	76.8	439	5 ABP25885	Abp25885 Streptoco
19	53	76.8	442	6 ABU02156	Abu02156 S. pneumo
20	53	76.8	442	8 ADK46665	Adk46665 Streptoco
21	53	76.8	443	3 AAY81621	Aay81621 Streptoco
22	53	76.8	458	8 ADR94508	Adr94508 Novel S.
23	53	76.8	458	9 AEA58378	Aea58378 Streptoco

#### ALIGNMENTS

RESULT 1  
AAO26386  
ID AAO26386 standard; peptide; 12 AA.  
XX  
AC AAO26386;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 6.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
(AKIV-) AKIVA LLC.  
XX  
O'daly JA;  
XX  
WPI; 2003-018763/01.  
XX  
New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
Claim 1; Page 43; 56pp; English.  
XX  
The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The sequences of immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention

ABO60156 Human gen  
Adr98843 Lung spec  
Abg23687 Novel hum  
Adr98840 Lung spec  
Adr98908 Lung spec  
Adb64789 Human pro  
Adr98947 Lung spec  
Adr98903 Lung spec  
Adr98906 Lung spec  
Adr98904 Lung spec  
Abm81872 Tumour-as  
Aea15473 Human pol  
Adr98902 Lung spec  
Adr98907 Lung spec  
Adr98901 Lung spec  
Adb75392 Prostate  
Adj76198 Marker ge  
Ads85099 Mouse ato  
Adj75405 Marker ge  
Adn04448 Antipsori  
Adn98876 Novel hum  
Adr98900 Lung spec

```
SQ Sequence 12 AA;
Query Match 100.0%; Score 69; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPTYNTIDDL 12
   |||||
DB 1 NYSPTYNTIDDL 12

RESULT 2
ADV04411
ID ADV04411 standard; peptide; 12 AA.
XX
AC ADV04411;
XX
DT 24-FEB-2005 (first entry)
XX
DE Leishmania immunotherapeutic peptide SEQ ID NO:6.
XX
KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;
KW psoriasis.
XX
OS Leishmania sp.
XX
PN US2004241168-A1.
XX
PD 02-DEC-2004.
XX
PF 09-FEB-2004; 2004US-00774928.
XX
PR 16-MAR-2001; 2001US-00809003.
PR 17-OCT-2003; 2003US-00687892.
XX
PA (ODAL/) ODALY J A.
XX
PI Odaly JA;
XX
WPI; 2005-011563/01.
XX
Inhibiting selectively T-cell rolling in human, by administering compound
PT interfering with cutaneous lymphocyte antigen-E selectin interaction and
PT leukocyte function associated antigen-1 interaction.
XX
Claim 9; SEQ ID NO 6; 21pp; English.
XX
The invention relates to a novel method for selectively inhibiting T-cell
CC rolling in a human host, comprising administering a compound that
CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E
CC selectin interaction and leukocyte function associated antigen (LFA)-
CC 1/intercellular adhesion molecule (ICAM) and very late antigen
CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound
CC includes an immunotherapeutic agent, which comprises a purified protein
CC extract that is isolated by diethylaminoethyl Sephadex chromatography of
CC a Nonidet P-40 insoluble particulate antigen fraction derived from
CC isolated killed cells of amastigotes from one or more species of the
CC Leishmania genus, where the particulate antigen fraction is solubilized
CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied
CC to diethylaminoethyl Sephadex and is eluted with a solution comprising
CC 0.1 M sodium chloride, 8 M urea and 0.025 M
CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract
CC includes polypeptides having apparent molecular weights of 73, 80 and 82
CC kDa, after total reduction and alkylation. The species is L. amazonensis,
CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the
CC invention has antipsoriatic and immunostimulant activity, and inhibits T-
CC cell rolling by interfering with CLA-E selectin interaction and LFA-
CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively
CC inhibiting T-cell rolling in a human host, and for treating psoriasis.
CC The present sequence represents a Leishmania peptide used in the
CC invention.
XX
SQ Sequence 12 AA;
Query Match 100.0%; Score 69; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPTYNTIDDL 12
   |||||
DB 1 NYSPTYNTIDDL 12

RESULT 3
AAU68343
ID AAU68343 standard; peptide; 13 AA.
XX
AC AAU68343;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human Breast cancer-associated protein isoform, BPI-342 peptide.
XX
KW Human; Breast cancer-associated protein isoform; breast cancer;
KW immunogen; cytostatic; BPI; tryptic digest peptide.
XX
OS Homo sapiens.
XX
PN WO200171357-A2.
XX
PD 27-SEP-2001.
XX
PF 20-MAR-2001; 2001WO-GB001219.
XX
PR 20-MAR-2000; 2000GB-00006695.
PR 24-MAR-2000; 2000GB-00007265.
XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI Herath HMAAC, O'hare MJ, Page MJ, Parekh RB, Waterfield MD;
XX
WPI; 2001-611532/70.
XX
Identifying proteins for clinical screening, diagnosis and prognosis of
PT breast cancer, comprises detecting Breast Cancer-Associated Protein
PT isoforms (BPis) using two-dimensional electrophoresis.
XX
Claim 9; Page 50; 197pp; English.
XX
The invention relates to diagnosing, determining the stage or severity,
CC or identifying the risk of a subject developing cancer (especially breast
CC cancer), or monitoring the effect of therapy on a subject with cancer,
CC comprising analysing a test sample using two-dimensional electrophoresis
CC and detecting Breast Cancer-Associated Protein Isoforms (BPis). The
CC methods disclosed are used for the diagnosis and prognosis of breast
CC cancer, for determining the severity of breast cancer, and for
CC identifying a subject at risk of developing breast cancer, and monitoring
CC the effect of therapy administered to a subject. Antibodies raised
CC against the binding domain of a BPI, the binding domain of a BPI, a
CC nucleic acid encoding a BPI, or a nucleic acid that inhibits the function
CC of a BPI can be incorporated into a pharmaceutical composition for
CC treating or preventing breast cancer. The methods use sensitive and
CC specific biomarkers provide early diagnosis of breast cancer, and the
CC compositions are more potent, specific, and has a more rapid effect with
CC fewer side effects than other prior art methods. The present sequence is
CC a tryptic digest peptide from a BPI of the invention
XX
SQ Sequence 13 AA;
Query Match 100.0%; Score 69; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPTYNTIDDL 12
   |||||
DB 1 NYSPTYNTIDDL 12
```



RESULT 4  
ID AEE36578 standard; peptide; 27 AA.  
XX AEE36578;  
XX AC  
XX DT 09-FEB-2006 (first entry)  
XX DE Human serum N-linked glycopeptide SEQ ID NO: 682.  
XX KW Bioinformatics; blood; serum; plasma protein; protein detection;  
XX mass spectroscopy; proteomics; glycosylation; diagnosis; cancer;  
XX KW rheumatoid arthritis; antiarthritic; antirheumatic; psychiatric disorder;  
XX KW neuroleptic; neurological disease; infection; antimicrobial.  
XX OS Homo sapiens.  
XX PN WO2005114221-A2.  
XX PD 01-DEC-2005.  
XX PF 20-MAY-2005; 2005WO-US017842.  
XX PR 21-MAY-2004; 2004US-0573593P.  
XX PA (SYST-) INST SYSTEMS BIOLOGY.  
XX PI Aebersold RH, Zhang H;  
XX WPI; 2006-020173/02.  
XX  
XX Identifying glycopolypeptides in a serum or plasma sample, by identifying  
XX released sample glycopeptide fragments that correspond to standard  
XX peptides.  
XX Claim 1; SEQ ID NO 682; 193pp; English.  
XX  
XX The invention relates to identifying glycopolypeptides in a serum or  
XX plasma sample comprising immobilizing derivatized sample  
XX glycopolypeptides to a solid support, releasing the sample glycopeptide  
XX fragments from the solid support, adding to the released sample  
XX glycopeptide fragments standard peptides, and identifying released sample  
XX glycopeptide fragments that correspond to standard peptides added by mass  
XX spectroscopy. Also included are a method for identifying one or more  
XX diagnostic markers for a disease, a composition comprising peptides  
XX containing the glycosylation sites (AEE35897-AEE39378, where the peptides  
XX each correspond to peptide fragments derived by cleavage of polypeptides  
XX using the same cleavage reagent) and a kit comprising peptides containing  
XX the glycosylation sites (AEE35897-AEE39378). The methods are useful for  
XX identifying glycopolypeptides in a serum or plasma sample. The methods  
XX can be used for blood serum profiling for the detection of prognostic and  
XX diagnostic protein markers. It can also be used to identify and/or  
XX validate drug targets and to evaluate drug efficacy, drug dosing, and/or  
XX drug toxicity. The methods can also be used for the detection of changes  
XX in the state of glycosylation of proteins based on the concurrent  
XX application of protein abundance measurement of protein glycosylation on  
XX the same sample. The method allows fast throughput and simplicity. It can  
XX be readily adapted for high throughput analysis of samples, which can be  
XX particularly advantageous for the analysis of clinical specimens. The  
XX method can also be automated to facilitate the processing of multiple  
XX samples. The present sequence is a human glycopeptide comprising an N-  
XX linked glycosylation site, suitable for use as a reference peptide in the  
XX method of the invention.  
XX  
XX Sequence 27 AA;

Query Match 100.0%; Score 69; DB 10; Length 27;  
Best Local Similarity 100.0%; Pred. No. 9.7e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPYNTIDDL 12  
DB |||||  
2 NYSPYNTIDDL 13  
RESULT 5  
ID ADS88365 standard; protein; 622 AA.  
XX ADS88365;  
XX DT 18-NOV-2004 (first entry)  
XX DE Human protein of a TNF-alpha signalling pathway protein complex Seq 220.  
XX KW protein complex; tumour necrosis factor-alpha signalling pathway;  
XX TNF-alpha; chronic inflammatory disease; rheumatoid arthritis;  
XX KW inflammatory bowel disease; infectious disease; septic shock;  
XX KW bacterial infection; neurological disease; stroke-induced inflammation;  
XX KW neurodegenerative disease; cancer; antiinflammatory; antiarthritic;  
XX KW antirheumatic; cytostatic; antibacterial; gene therapy; human.  
XX OS Homo sapiens.  
XX PN WO2004035783-A2.  
XX PD 29-APR-2004.  
XX PF 24-SEP-2003; 2003WO-EP050655.  
XX PR 26-SEP-2002; 2002EP-00021809.  
XX PR 10-FEB-2003; 2003EP-00100274.  
XX (CELL-) CELLZOME AG.  
XX Bouwmeester T, Huhse B, Bauch A, Ruffner H, Bauer A, Kuester B;  
XX Superti-Furga G, Kruse U;  
XX WPI; 2004-348460/32.  
XX  
XX New protein complex comprising at least one first and second protein of  
XX the Tumor Necrosis Factor-alpha(TNF-alpha)-signalling pathway, useful for  
XX diagnosing or treating inflammation, neurological diseases, infectious  
XX diseases or cancer.  
XX Example; SEQ ID NO 220; 1980pp; English.  
XX  
XX This invention relates to novel protein complexes of the tumour necrosis  
XX factor-alpha (TNF-alpha) signalling pathway. Specifically, it refers to  
XX methods for preparing these complexes comprising at least two component  
XX proteins, as well as screening methods to identify modulators of the  
XX pathway, which include antibodies, agonists and antagonists thereof. The  
XX present invention describes a protein complex and kit that are useful for  
XX diagnosing, prognosing or treating chronic inflammatory diseases such as  
XX rheumatoid arthritis and inflammatory bowel disease; infectious diseases  
XX such as septic shock and bacterial infections; neurological diseases such  
XX as stroke-induced inflammation in neurons; neurodegenerative diseases and  
XX cancer. Accordingly, these complexes can be used for the development of  
XX pharmaceutical compositions that exhibit antiinflammatory, antiarthritic,  
XX antirheumatic, cytostatic and antibacterial activities and can be used  
XX for gene therapy purposes. In particular, the invention further provides  
XX siRNA-oligonucleotides useful for inhibiting protein expression for in  
XX vitro or cell culture assays. This polypeptide is a human protein that  
XX can be used in combination with other proteins provided in the  
XX specification to form novel complexes of the TNF-alpha signalling pathway  
XX of the invention.  
XX  
XX Sequence 622 AA;

Query Match 100.0%; Score 69; DB 8; Length 622;  
Best Local Similarity 100.0%; Pred. No. 0.0041;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NYSPPYNTIDDL 12  
 Db 199 NYSPPYNTIDDL 210

RESULT 6  
 AEA15476  
 ID AEA15476 standard; protein; 622 AA.  
 AC AEA15476;  
 XX  
 DT 28-JUL-2005 (first entry)  
 DE Human polypeptide #95.  
 XX  
 KW Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
 KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
 KW cytostatic; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005047534-A2.  
 XX  
 PD 26-MAY-2005.  
 XX  
 PF 15-OCT-2004; 2004WO-EP011599.  
 XX  
 PR 28-OCT-2003; 2003EP-00024565.  
 XX  
 PA (FARB ) BAYER HEALTHCARE AG.  
 XX  
 PI Wirtz R, Munnes M;  
 DR WPI; 2005-372393/38.  
 DR N-PSDB; AEA15398.  
 XX

PT Predicting a response to cancer treatment by detecting at least 2  
 PT markers, which are genes or genomic nucleic acid sequences that are  
 PT located on one chromosomal region, which is altered in malignant  
 PT neoplasia.  
 XX

PS Claim 7; SEQ ID NO 438; 464pp; English.  
 XX

CC The invention relates to a method of predicting response to cancer  
 CC treatment comprising detection of at least 2 markers, where the markers  
 CC are genes and fragments or genomic nucleic acid sequences that are  
 CC located on one chromosomal region, which is altered in malignant  
 CC neoplasia. The invention also relates to a method for the prediction,  
 CC diagnosis or prognosis of malignant neoplasia, methods for detecting  
 CC deregulations in malignant neoplasia and breast cancer, a method of  
 CC determining the phenotype of a cell or tissue, a method for identifying  
 CC genomic regions which are altered on the chromosomal level and encode  
 CC genes that are linked by function and are differentially expressed in  
 CC malignant neoplasia and breast cancer, methods of screening for agents  
 CC which regulate the activity of a polypeptide or a polynucleotide and  
 CC antibodies that specifically bind to a full length or partial  
 CC polypeptide. The method is useful for predicting response to cancer  
 CC treatment. The methods and compositions are useful for predicting,  
 CC diagnosing, prognosing, preventing or treating malignant neoplasia  
 CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
 CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
 CC lung cancer. This sequence represents a human polypeptide used in the  
 CC scope of the invention.  
 XX

SQ Sequence 622 AA;  
 Query Match 100.0%; Score 69; DB 9; Length 622;  
 Best Local Similarity 100.0%; Pred. No. 0.0041;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NYSPPYNTIDDL 12  
 Db 199 NYSPPYNTIDDL 210

RESULT 7  
 AEB13340  
 ID AEB13340 standard; protein; 622 AA.  
 XX  
 AC AEB13340;  
 XX  
 DT 08-SEP-2005 (first entry)  
 XX  
 DE Mutant keratin-9 SEQ ID NO 5.  
 XX  
 KW antigen; tumor-associated antigen; diagnostic; therapeutic; tumor marker;  
 KW squamous cell carcinoma;  
 KW autoantibody-mediated identification of antigens; AMIDA; cytostatic;  
 KW neoplasm; cancer; skin tumor; ear, nose, throat disease; mutant;  
 KW keratin-9.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO2005061537-A2.  
 XX  
 PD 07-JUL-2005.  
 XX  
 PF 22-DEC-2004; 2004WO-EP014634.  
 XX  
 PR 22-DEC-2003; 2003DE-01060456.  
 XX  
 PA (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.  
 PA (VAEC-) VAECGENE BIOTECH GMBH.  
 XX  
 PI Gires O, Zeidler R, Muenz M, Schaffrik M;  
 DR WPI; 2005-479443/48.  
 XX

PT Antigen for skin tumors in the nose and throat, uses at least one protein  
 PT from a selected sequence or their variants with a hybridized isolated  
 PT nucleic acid encoding the transcribed product.  
 XX

PS Claim 1; SEQ ID NO 5; 54pp; German.  
 XX

CC This invention describes the detection of novel antigens associated with  
 CC squamous cell carcinomas using autoantibody-mediated identification of  
 CC antigens (AMIDA). The invention also describes methods for producing  
 CC antigen-presenting cells and T cells that are specific for the antigens  
 CC described. The antigens can be used in a diagnostic formulation or as a  
 CC pharmaceutical product for therapy or as a vaccine, for diagnosis and  
 CC therapy of squamous cell tumors and skin tumors associated with the ear,  
 CC nose and throat (ENT). The tumor antigen is an improved diagnostic marker  
 CC for tumor identification, and is also an effective therapeutic agent.  
 CC This sequence represents the tumor-associated antigen corresponding to a  
 CC mutant keratin-9.  
 XX

SQ Sequence 622 AA;  
 Query Match 100.0%; Score 69; DB 9; Length 622;  
 Best Local Similarity 100.0%; Pred. No. 0.0041;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NYSPPYNTIDDL 12  
 Db 199 NYSPPYNTIDDL 210

RESULT 8  
 ADF55631  
 ID ADF55631 standard; protein; 600 AA.  
 XX  
 AC ADF55631;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX

DE Protein (SEQ ID 1) related to oyster pearl keratin protein.  
 XX  
 KW pearl protein composite; PPC; mother of pearl manufacture; jewellery;  
 XW nanotechnology; oyster; keratin.  
 XX  
 OS Unidentified.  
 XX  
 PN JP2003012696-A.  
 XX  
 PD 15-JAN-2003.  
 XX  
 PF 03-JUL-2001; 2001JP-00202033.  
 XX  
 PR 03-JUL-2001; 2001JP-00202033.  
 XX  
 PA (MATSU) MATSUSHIRO A.  
 XX  
 DR WPI; 2003-451748/43.  
 XX  
 XX Pearl protein composite useful in manufacture of mother of pearl,  
 PT comprises at least two protein components having preset molecular weight  
 PT as determined by sodium dodecyl sulfate gel electrophoresis.  
 XX  
 XX Example 4; SEQ ID NO 1; 25pp; Japanese.  
 PS  
 CC The invention relates to a novel pearl protein composite (PPC) comprising  
 CC at least two protein components as determined by sodium dodecyl sulphate  
 CC (SDS) gel electrophoresis. The composite of the invention may be useful  
 CC for the manufacture of mother of pearl for jewellery and also in  
 CC nanotechnology. The pearl protein composite (PPC) is effectively utilised  
 CC in manufacture of artificial pearls. The current sequence is that of the  
 CC protein (SEQ ID 1) of the invention which is related to oyster pearl  
 CC keratin protein and is homologous to a human keratin.  
 XX  
 XX Sequence 600 AA;  
 SQ

Query Match 94.2%; Score 65; DB 7; Length 600;  
 Best Local Similarity 100.0%; Pred. No. 0.019;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYSPYNTTIDD 11  
 |||||  
 DB 199 NYSPYNTTIDD 209

RESULT 9  
 AAW55075  
 ID AAW55075 standard; protein; 420 AA.  
 XX  
 AC AAW55075;  
 XX  
 DT 02-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae SP0016 protein.  
 XX  
 KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 XW detection; pneumonia; otitis media; meningitis.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9818930-A2.  
 XX  
 PD 07-MAY-1998.  
 XX  
 PF 30-OCT-1997; 97WO-US019422.  
 XX  
 PR 31-OCT-1996; 96US-0029960P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;  
 XX WPI; 1998-272224/24.  
 DR

DR N-PSDB; AAV27336.  
 XX  
 PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae  
 PT - or their epitope-containing fragments, useful in protective or  
 PT therapeutic vaccines, and for diagnosis.  
 XX  
 PS Claim 11; Page 54; 118pp; English.  
 XX  
 CC The present sequence represents a protein from Streptococcus pneumoniae.  
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose  
 XX  
 SQ Sequence 420 AA;  
 CC

Query Match 76.8%; Score 53; DB 2; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YSPYNTTID 10  
 |||||  
 DB 368 YSPYNTTID 376

RESULT 10  
 ABP54569  
 ID ABP54569 standard; protein; 420 AA.  
 XX  
 AC ABP54569;  
 XX  
 DT 04-SEP-2002 (first entry)  
 XX  
 DE S. pneumoniae SP016 protein sequence SEQ ID NO:26.  
 XX  
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 XW antibacterial; Streptococcal infection; detection.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN US2002061545-A1.  
 XX  
 PD 23-MAY-2002.  
 XX  
 PF 22-JAN-2001; 2001US-00765272.  
 XX  
 PR 30-OCT-1997; 97US-00961083.  
 XX  
 PA (CHOI/) CHOI G H.  
 PA (KUNS/) KUNSCH C A.  
 PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.  
 PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 DR WPI; 2002-479261/51.  
 DR N-PSDB; ABQ84804.  
 XX  
 XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus  
 PT and for preventing or attenuating disease caused by Streptococcus

PT infection.  
PS Claim 11; Page 23; 70pp; English.  
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the  
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.  
CC pneumoniae antigens have antibacterial activity and can be used in  
CC vaccines. The S. pneumoniae antigens can also be used to prevent or  
CC attenuate a Streptococcal infection in an animal. The polynucleotides  
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus  
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning  
CC of S. pneumoniae ORFs (open reading frames) which are used in an example  
CC from the present invention  
XX  
SQ Sequence 420 AA;  
  
Query Match 76.8%; Score 53; DB 5; Length 420;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 YSPYNTID 10  
Db 368 YSPYNTID 376  
|||||  
  
RESULT 11  
ADC45107  
ID ADC45107 standard; protein; 420 AA.  
XX  
AC ADC45107;  
XX  
XX 18-DEC-2003 (first entry)  
XX  
XX S. pneumoniae antigenic protein SP016.  
XX  
KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX US6573082-B1.  
XX  
XX 03-JUN-2003.  
XX  
XX 28-MAR-2000; 2000US-00536784.  
XX  
PR 31-OCT-1996; 96US-0029960P.  
PR 30-OCT-1997; 97US-00961083.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
PI Rosen CA;  
XX  
XX WPI; 2003-764574/72.  
DR N-PSDB; ADC45106.  
XX  
XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
PT useful for producing vaccines for prevention or attenuation of infection  
PT by Streptococcus pneumoniae.  
XX  
XX Example 1; SEQ ID NO 26; 58pp; English.  
XX  
XX The invention relates to an isolated polynucleotide consisting of a  
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding  
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
CC antigens. Also included are making a recombinant vector by inserting the  
CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a  
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
CC antigen nucleic acids are useful as probes for use in diagnostic methods  
CC for detecting S. pneumoniae gene expression. The present sequence

CC represents an S. pneumoniae antigenic protein.  
XX  
SQ Sequence 420 AA;  
  
Query Match 76.8%; Score 53; DB 7; Length 420;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 YSPYNTID 10  
Db 368 YSPYNTID 376  
|||||  
  
RESULT 12  
RAY91342  
ID AAY91342 standard; protein; 438 AA.  
XX  
AC AAY91342;  
XX  
XX 30-MAY-2000 (first entry)  
XX  
XX Group B Streptococcus protein sequence SEQ ID NO:75.  
XX  
KW Group B Streptococcus; Streptococcus agalactiae; protein antigen;  
KW vaccine; screening; immunogen; detection; diagnosis; infection; antibody;  
KW antibody; antibacterial.  
XX  
OS Streptococcus agalactiae.  
XX  
XX WO200006736-A2.  
XX  
XX 10-FEB-2000.  
XX  
XX 27-JUL-1999; 99WO-GB002444.  
XX  
PR 27-JUL-1998; 98GB-00016335.  
PR 19-MAR-1999; 99US-0125163P.  
XX  
XX (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
XX Le Page RWF, Wells JM, Hanniffy SB;  
PI WPI; 2000-195299/17.  
XX  
XX New Group B Streptococcus protein, useful as vaccine, for diagnosis of  
PT Streptococcal infections and for screening of antibodies or affibodies.  
XX  
XX Claim 1; Fig 1; 123pp; English.  
XX  
XX AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given in  
CC AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also  
CC known as Streptococcus agalactiae. The GBS polynucleotides and  
CC polypeptides have antibacterial activity. Immunogenic compositions  
CC comprising GBS polynucleotides or polypeptides can be used as vaccines  
CC and for the treatment or prophylaxis of GBS infection. The  
CC polynucleotides and polypeptides can also be used in the detection of GBS  
CC and for screening DNA encoding bacterial cell envelope associated or  
CC secreted antigens in gram positive bacteria. AAA05873 to AAA05941  
CC represent primers used in the exemplification of the present invention  
XX  
XX Sequence 438 AA;  
  
Query Match 76.8%; Score 53; DB 3; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 YSPYNTID 10  
Db 387 YSPYNTID 395  
|||||  
  
RESULT 13  
ABP25884

ID ABP25884 standard; protein; 438 AA.  
XX AC ABP25884;  
XX DT 02-JUL-2002 (first entry)  
XX DE Streptococcus polypeptide SEQ ID NO 944.  
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX OS Streptococcus agalactiae.  
XX PN WO200234771-A2.  
XX PD 02-MAY-2002.  
XX PF 29-OCT-2001; 2001WO-GB004789.  
XX PR 27-OCT-2000; 2000GB-00026333.  
XX PR 24-NOV-2000; 2000GB-00028727.  
XX PR 07-MAR-2001; 2001GB-00005640.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
XX PI Tettelin H;  
XX DR WPI; 2002-352536/38.  
XX DR N-PSDB; ABN66515.  
XX PT New Streptococcus protein for the treatment or prevention of infection or  
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for  
XX PT detecting a compound that binds to the protein.  
XX PS Claim 1; Page 3248; 4525pp; English.  
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
XX CC the specification. The proteins have antibacterial and antiinflammatory  
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX CC antibodies that bind (I) are used in the manufacture of medicaments for  
XX CC the treatment or prevention of infection or disease caused by  
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX CC biological sample. (I) is used to determine whether a compound binds to  
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be  
XX CC used in gene therapy. Antibodies to (I) are used for affinity  
XX CC chromatography, immunoassays, and distinguishing/identifying  
XX CC Streptococcus proteins  
XX SQ Sequence 438 AA;  
Query Match 76.8%; Score 53; DB 5; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 YSPYNTID 10  
Db 387 YSPYNTID 395  
RESULT 15  
ADV89335  
ID ADV89335 standard; protein; 438 AA.  
XX AC ADV89335;  
XX DT 24-FEB-2005 (first entry)  
XX AC ABP29681 standard; protein; 438 AA.  
XX AC ABP29681;  
ID ABP25884 standard; protein; 438 AA.  
XX AC ABP25884;  
XX DT 02-JUL-2002 (first entry)  
XX DE Streptococcus polypeptide SEQ ID NO 944.  
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX OS Streptococcus agalactiae.  
XX PN WO200234771-A2.  
XX PD 02-MAY-2002.  
XX PF 29-OCT-2001; 2001WO-GB004789.  
XX PR 27-OCT-2000; 2000GB-00026333.  
XX PR 24-NOV-2000; 2000GB-00028727.  
XX PR 07-MAR-2001; 2001GB-00005640.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
XX PI Tettelin H;  
XX DR WPI; 2002-352536/38.  
XX DR N-PSDB; ABN66515.  
XX PT New Streptococcus protein for the treatment or prevention of infection or  
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for  
XX PT detecting a compound that binds to the protein.  
XX PS Claim 1; Page 3248; 4525pp; English.  
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
XX CC the specification. The proteins have antibacterial and antiinflammatory  
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX CC antibodies that bind (I) are used in the manufacture of medicaments for  
XX CC the treatment or prevention of infection or disease caused by  
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX CC biological sample. (I) is used to determine whether a compound binds to  
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be  
XX CC used in gene therapy. Antibodies to (I) are used for affinity  
XX CC chromatography, immunoassays, and distinguishing/identifying  
XX CC Streptococcus proteins  
XX SQ Sequence 438 AA;  
Query Match 76.8%; Score 53; DB 5; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 YSPYNTID 10  
Db 387 YSPYNTID 395  
RESULT 14  
ABP29681  
ID ABP29681 standard; protein; 438 AA.  
XX AC ABP29681;  
ID ABP25884 standard; protein; 438 AA.  
XX AC ABP25884;  
XX DT 02-JUL-2002 (first entry)  
XX DE Streptococcus polypeptide SEQ ID NO 8538.  
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX OS Streptococcus agalactiae.  
XX PN WO200234771-A2.  
XX PD 02-MAY-2002.  
XX PF 29-OCT-2001; 2001WO-GB004789.  
XX PR 27-OCT-2000; 2000GB-00026333.  
XX PR 24-NOV-2000; 2000GB-00028727.  
XX PR 07-MAR-2001; 2001GB-00005640.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
XX PI Tettelin H;  
XX DR WPI; 2002-352536/38.  
XX DR N-PSDB; ABN70312.  
XX PT New Streptococcus protein for the treatment or prevention of infection or  
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for  
XX PT detecting a compound that binds to the protein.  
XX PS Claim 1; Page 3964; 4525pp; English.  
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
XX CC the specification. The proteins have antibacterial and antiinflammatory  
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX CC antibodies that bind (I) are used in the manufacture of medicaments for  
XX CC the treatment or prevention of infection or disease caused by  
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX CC biological sample. (I) is used to determine whether a compound binds to  
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be  
XX CC used in gene therapy. Antibodies to (I) are used for affinity  
XX CC chromatography, immunoassays, and distinguishing/identifying  
XX CC Streptococcus proteins  
XX SQ Sequence 438 AA;  
Query Match 76.8%; Score 53; DB 5; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 YSPYNTID 10  
Db 387 YSPYNTID 395  
RESULT 15  
ADV89335  
ID ADV89335 standard; protein; 438 AA.  
XX AC ADV89335;  
XX DT 24-FEB-2005 (first entry)  
XX AC ABP29681 standard; protein; 438 AA.  
XX AC ABP29681;  
ID ABP25884 standard; protein; 438 AA.  
XX AC ABP25884;  
XX DT 02-JUL-2002 (first entry)  
XX DE Streptococcus polypeptide SEQ ID NO 8538.  
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX OS Streptococcus agalactiae.  
XX PN WO200234771-A2.  
XX PD 02-MAY-2002.  
XX PF 29-OCT-2001; 2001WO-GB004789.  
XX PR 27-OCT-2000; 2000GB-00026333.  
XX PR 24-NOV-2000; 2000GB-00028727.  
XX PR 07-MAR-2001; 2001GB-00005640.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
XX PI Tettelin H;  
XX DR WPI; 2002-352536/38.  
XX DR N-PSDB; ABN70312.  
XX PT New Streptococcus protein for the treatment or prevention of infection or  
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for  
XX PT detecting a compound that binds to the protein.  
XX PS Claim 1; Page 3964; 4525pp; English.  
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
XX CC the specification. The proteins have antibacterial and antiinflammatory  
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX CC antibodies that bind (I) are used in the manufacture of medicaments for  
XX CC the treatment or prevention of infection or disease caused by  
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX CC biological sample. (I) is used to determine whether a compound binds to  
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be  
XX CC used in gene therapy. Antibodies to (I) are used for affinity  
XX CC chromatography, immunoassays, and distinguishing/identifying  
XX CC Streptococcus proteins  
XX SQ Sequence 438 AA;

DE	Streptococcus agalactiae protein sequence, SEQ ID 1729.
XX	
XX	Antibacterial; Vaccines; bacterial infection.
XX	
OS	Streptococcus agalactiae.
XX	
PN	FR2824074-A1.
XX	
PD	31-OCT-2002.
XX	
XX	26-APR-2001; 2001FR-00005642.
XX	
PX	26-APR-2001; 2001FR-00005642.
PR	(INSP ) INST PASTEUR.
XX	(CNRS ) CNRS CENT NAT RECH SCI.
PA	
PI	Glaser P, Rusniok C, Chevalier F, Frangeul L, Laloui L;
PP	Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;
XX	
DR	WPI; 2004-101891/11.
XX	
XX	Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT	agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT	and identification of therapeutic targets.
XX	
PS	Claim 6; SEQ ID NO 1729; 2687pp; French.
XX	
CC	The present invention relates to novel Streptococcus agalactiae
CC	nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
CC	ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
CC	agalactiae involved in the synthesis of amino acids, cell membranes,
CC	intermediate (central) metabolism, energetic metabolism including purines,
CC	phospholipid metabolism, nucleotide metabolism including fatty acid and
CC	pyrimidines and/or nucleosides, regulatory functions, replication,
CC	transcription, translation, protein transport, adaptation to atypical
CC	conditions, sensitivity to medicines and/or analogues, functions related
CC	to transposons, biosynthesis of cofactors, prosthetic groups and
CC	transporters, cell membrane proteins and cellular machinery. (I) are
CC	useful for the detection and/or amplification of nucleic acids.
CC	Pharmaceutical composition comprising (I) or (II) are useful for
CC	treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is
CC	equivalent for the present basic patent FR2824074A1. WO200292818A2
CC	contains 6617 sequence whereas the present patent only contains 2344
CC	sequences.
XX	
SQ	Sequence 438 AA;
	Query Match 76.8%; Score 53; DB 8; Length 438;
	Best Local Similarity 100.0%; Pred No. 1.7;
	Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps
Oy	2 YSPYNTID 10 
Db	387 YSPYNTID 395

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 87.8447 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-6

Perfect score: 69

Sequence: 1 NYSPTYNTIDDL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	100.0	12	5	US-10-774-928-6
2	69	100.0	13	3	US-09-988-493-239
3	53	76.8	420	3	US-09-765-272-26
4	53	76.8	420	6	US-11-106-649-26
5	53	76.8	438	3	US-09-769-736-137
6	53	76.8	438	6	US-11-123-241-137
7	53	76.8	439	3	US-10-474-792-96
8	53	76.8	442	3	US-09-769-787-121
9	53	76.8	442	5	US-10-472-928-3466
10	53	76.8	442	5	US-10-873-528-121
11	53	76.8	458	5	US-10-617-320-3443
12	50	72.5	101	4	US-10-425-115-308578
13	50	72.5	166	4	US-10-029-386-33790
14	50	72.5	221	5	US-10-450-763-54046
15	50	72.5	355	4	US-10-104-047-2943
16	50	72.5	355	6	US-11-072-512-2943
17	50	72.5	458	4	US-10-205-823-216
18	50	72.5	458	6	US-11-051-454-216
19	50	72.5	469	5	US-10-631-467-1450
20	50	72.5	470	5	US-10-631-467-657
21	50	72.5	472	6	US-11-037-713-35
22	50	72.5	558	5	US-10-450-763-54047
23	50	72.5	567	5	US-10-450-763-54049
24	50	72.5	593	4	US-10-435-696-50
25	50	72.5	618	3	US-09-925-300-1381
26	50	72.5	887	5	US-10-450-763-54043
27	50	72.5	1113	5	US-10-450-763-37061

28 45 65.2 324 5 US-10-501-282-172 Sequence 172, App  
29 45 65.2 352 5 US-10-501-282-174 Sequence 174, App  
30 45 65.2 381 5 US-10-501-282-176 Sequence 176, App  
31 45 65.2 392 5 US-10-501-282-178 Sequence 178, App  
32 44 63.8 1167 6 US-10-335-977-8633 Sequence 8633, App  
33 44 63.8 1167 6 US-11-052-554A-121 Sequence 121, App  
34 44 63.8 1178 4 US-10-335-977-8634 Sequence 8634, App  
35 44 63.8 1183 4 US-10-335-977-8635 Sequence 8635, App  
36 43 62.3 13 3 US-09-988-493-48 Sequence 48, Appl  
37 43 62.3 49 4 US-10-767-701-50836 Sequence 50836, A  
38 43 62.3 142 3 US-09-738-973-213 Sequence 213, App  
39 43 62.3 142 3 US-09-854-133-213 Sequence 213, App  
40 43 62.3 142 4 US-10-144-649A-213 Sequence 213, App  
41 43 62.3 143 3 US-09-925-301-1456 Sequence 1456, App  
42 43 62.3 157 5 US-10-450-763-40381 Sequence 40381, A  
43 43 62.3 247 4 US-10-104-047-3859 Sequence 3859, App  
44 43 62.3 247 6 US-11-072-512-3859 Sequence 3859, App  
45 43 62.3 354 5 US-10-732-923-15525 Sequence 15525, A

#### ALIGNMENTS

##### RESULT 1

US-10-774-928-6  
; Sequence 6, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-6

Query Match 100.0%; Score 69; DB 5; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.00022; Mismatches 0; Indels 0; Gaps 0;  
Matches 12; Conservative 0;

QY 1 NYSPTYNTIDDL 12  
Db 1 NYSPTYNTIDDL 12

##### RESULT 2

US-09-988-493-239  
; Sequence 239, Application US/09988493  
; Publication No. US20030064419A1  
; GENERAL INFORMATION:  
; APPLICANT: Herath, Herath Mudiyanseelage Athula Chandrasiri  
; APPLICANT: O'Hare, Michael John  
; APPLICANT: Page, Martin John  
; APPLICANT: Parekh, Rajesh Bhikhu  
; APPLICANT: Waterfield, Michael Derek  
; TITLE OF INVENTION: Proteins, Genes, and Their Use for  
; TITLE OF INVENTION: Diagnosis and Treatment of Breast Cancer  
; FILE REFERENCE: 2543-1-024  
; CURRENT APPLICATION NUMBER: US/09/988,493  
; CURRENT FILING DATE: 2002-05-21  
; PRIOR APPLICATION NUMBER: PCT/GB01/01219  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: GB 0006695.1  
; PRIOR FILING DATE: 2000-03-20  
; PRIOR APPLICATION NUMBER: GB 0007265.2  
; PRIOR FILING DATE: 2000-02-24  
; NUMBER OF SEQ ID NOS: 308

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 239
; LENGTH: 13
; TYPE: PRT
; ORGANISM: homo sapien
US-09-988-493-239

Query Match      100.0%; Score 69; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 NYSPTYNTIDDL 12
        |||||
Db      1 NYSPTYNTIDDL 12

RESULT 3
US-09-765-272-26
; Sequence 26, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:

US-09-765-272-26

Query Match      76.8%; Score 53; DB 3; Length 420;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 YSPYNTID 10
        |||||
Db      368 YSPYNTID 376

RESULT 4
US-11-106-649-26
; Sequence 26, Application US/11106649
; Publication No. US20050181439A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-26

Query Match      76.8%; Score 53; DB 6; Length 420;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 YSPYNTID 10
        |||||
Db      368 YSPYNTID 376

RESULT 5
US-09-769-736-137
; Sequence 137, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 137
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-137

Query Match      76.8%; Score 53; DB 3; Length 438;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 YSPYNTID 10
        |||||
Db      387 YSPYNTID 395

RESULT 6
US-11-123-241-137
; Sequence 137, Application US/11123241
; Publication No. US20060078565A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
```





```
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-121

Query Match      76.8%; Score 53; DB 5; Length 442;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YSPYNTID 10
Db 390 YSPYNTID 398

RESULT 11
US-10-617-320-3143
; Sequence 3143, Application US/10617320
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617.320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...458
; SEQUENCE DESCRIPTION: SEQ ID NO: 3143:
US-10-617-320-3143

Query Match      76.8%; Score 53; DB 5; Length 458;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YSPYNTID 10
```

```
|||||
Db 406 YSPYNTID 414

RESULT 12
US-10-425-115-308578
; Sequence 308578, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425.115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 308578
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_44492C.1.pep
US-10-425-115-308578

Query Match      72.5%; Score 50; DB 4; Length 101;
Best Local Similarity 66.7%; Pred. No. 2.4;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NYSPTYNTIDDL 12
Db 45 DYSPYFKTIEDL 56

RESULT 13
US-10-029-386-33790
; Sequence 33790, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029.386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33790
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL353997.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.98
; OTHER INFORMATION: SWISSPROT HIT: P08779, EVALUE 3.00e-27
US-10-029-386-33790

Query Match      72.5%; Score 50; DB 4; Length 166;
Best Local Similarity 66.7%; Pred. No. 4;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NYSPTYNTIDDL 12
Db 152 DYSPYFKTIEDL 163

RESULT 14
US-10-450-763-54046
```

```

; Sequence 54046, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; PRIOR FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 54046
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (128)...(143)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226A, p-value=1.857e-15, raw score of 12.77
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (128)...(221)
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,
; OTHER INFORMATION: accession name filament, E-value=1.4e-49, Pfam score of 178.1
US-10-450-763-54046

```

```

Query Match      72.5%; Score 50; DB 5; Length 221;
Best Local Similarity 66.7%; Pred. No. 5.4;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 NYSPPYNTIDDL 12
        :||||:|
DB      175 DYSPYKTIEDL 186

```

```

RESULT 15
US-10-104-047-2943
; Sequence 2943, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2943
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2943

```

```

Query Match      72.5%; Score 50; DB 4; Length 355;
Best Local Similarity 66.7%; Pred. No. 8.9;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 NYSPPYNTIDDL 12
        :||||:|
DB      139 DYSPYKTIEDL 150

```

Search completed: July 5, 2006, 20:52:00  
Job time : 87.8447 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 6.87379 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-6  
Perfect score: 69  
Sequence: 1 NYSPTYNTIDDL 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB pep.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	100.0	27	7	US-11-134-871-682
2	43	62.3	400	7	US-11-105-233-182
3	39	56.5	23	7	US-11-134-871-701
4	39	56.5	23	7	US-11-134-871-3401
5	39	56.5	540	6	US-10-449-902-51440
6	38	55.1	212	6	US-10-449-902-36492
7	38	55.1	212	6	US-10-449-902-51948
8	37	53.6	159	6	US-10-953-349-14759
9	37	53.6	315	6	US-10-953-349-28574
10	37	53.6	360	6	US-10-953-349-28573
11	37	53.6	394	6	US-10-953-349-28572
12	37	53.6	399	6	US-10-449-902-31745
13	37	53.6	399	6	US-10-449-902-55264
14	37	53.6	1051	6	US-10-449-902-55175
15	37	53.6	1466	6	US-10-985-570-1
16	37	53.6	1466	7	US-11-105-233-44
17	37	53.6	1466	7	US-11-105-233-167
18	36	52.2	25	7	US-11-134-871-2867
19	36	52.2	261	6	US-10-471-571A-496
20	36	52.2	424	7	US-11-105-233-183
21	36	52.2	599	6	US-10-471-571A-484
22	36	52.2	809	6	US-10-449-902-42942
23	35.5	51.4	315	6	US-10-953-349-9204
24	35.5	51.4	400	6	US-10-953-349-9203
25	35.5	51.4	401	6	US-10-953-349-9202

Sequence 52, Appl  
Sequence 2032, Ap  
Sequence 30011, A  
Sequence 42693, A  
Sequence 30010, A  
Sequence 2031, Ap  
Sequence 324, App  
Sequence 2030, Ap  
Sequence 5594, Ap  
Sequence 12, Appl  
Sequence 61, Appl  
Sequence 4, Appl  
Sequence 8, Appl  
Sequence 6, Appl  
Sequence 47769, A  
Sequence 10, Appl  
Sequence 40557, A  
Sequence 52546, A  
Sequence 55424, A  
Sequence 29415, A

## ALIGNMENTS

RESULT 1  
US-11-134-871-682  
; Sequence 682, Application US/11134871  
; Publication No. US20060141528A1  
; GENERAL INFORMATION:  
; APPLICANT: Abersold, Rudolf H.  
; APPLICANT: Zhang, Hui  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Quatification of Serum Glycoproteins  
; FILE REFERENCE: 66661-116  
; CURRENT APPLICATION NUMBER: US/11/134,871  
; CURRENT FILING DATE: 2005-05-20  
; PRIOR APPLICATION NUMBER: 60/573,593  
; PRIOR FILING DATE: 2004-05-21  
; NUMBER OF SEQ ID NOS: 3602  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 682  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-134-871-682

Query Match 100.0%; Score 69; DB 7; Length 27;  
Best Local Similarity 100.0%; Pred. No. 7.9e-06;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPTYNTIDDL 12  
|||||  
Db 2 NYSPTYNTIDDL 13

RESULT 2  
US-11-105-233-182  
; Sequence 182, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Thiagalingam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; TITLE OF INVENTION: Tumors  
; FILE REFERENCE: 1657/2001  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; CURRENT FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 182  
; LENGTH: 400  
; TYPE: PRT  
; ORGANISM: Homo sapiens

## US-11-105-233-182

Query Match 62.3%; Score 43; DB 7; Length 400;  
Best Local Similarity 66.7%; Pred. No. 4.1;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NVSPYNTIDDL 12  
Db 126 DYSHYTTIQDL 137

## RESULT 3

US-11-134-871-701  
; Sequence 701, Application US/111134871  
; Publication No. US20060141528A1  
; GENERAL INFORMATION:  
; APPLICANT: Aebersold, Rudolf H.  
; APPLICANT: Zhang, Hui  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Quantification of Serum Glycoproteins  
; FILE REFERENCE: 66661-116  
; CURRENT APPLICATION NUMBER: US/11/134,871  
; CURRENT FILING DATE: 2005-05-20  
; PRIOR APPLICATION NUMBER: 60/573,593  
; PRIOR FILING DATE: 2004-05-21  
; NUMBER OF SEQ ID NOS: 3602  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 701  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-134-871-701

Query Match 56.5%; Score 39; DB 7; Length 23;  
Best Local Similarity 60.0%; Pred. No. 0.93;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SPYNTIDDL 12  
Db 2 SPYNSDEI 11

## RESULT 4

US-11-134-871-3401  
; Sequence 3401, Application US/11134871  
; Publication No. US20060141528A1  
; GENERAL INFORMATION:  
; APPLICANT: Aebersold, Rudolf H.  
; APPLICANT: Zhang, Hui  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Quantification of Serum Glycoproteins  
; FILE REFERENCE: 66661-116  
; CURRENT APPLICATION NUMBER: US/11/134,871  
; CURRENT FILING DATE: 2005-05-20  
; PRIOR APPLICATION NUMBER: 60/573,593  
; PRIOR FILING DATE: 2004-05-21  
; NUMBER OF SEQ ID NOS: 3602  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3401  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-134-871-3401

Query Match 56.5%; Score 39; DB 7; Length 23;  
Best Local Similarity 60.0%; Pred. No. 0.93;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SPYNTIDDL 12  
Db 2 SPYNSDEI 11

## RESULT 5

US-10-449-902-51440  
; Sequence 51440, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51440  
; LENGTH: 540  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-51440

Query Match 56.5%; Score 39; DB 6; Length 540;  
Best Local Similarity 75.0%; Pred. No. 27;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 YYNTIDDL 12  
Db 420 YYNTIDM 427

## RESULT 6

US-10-449-902-36492  
; Sequence 36492, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 36492  
; LENGTH: 212  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-36492

Query Match 55.1%; Score 38; DB 6; Length 212;  
Best Local Similarity 60.0%; Pred. No. 15;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NVSPYNTID 10  
Db 106 NKSPYHEIVD 115

## RESULT 7

US-10-449-902-51948  
; Sequence 51948, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:

; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51948  
; LENGTH: 212  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-51948

Query Match 55.1%; Score 38; DB 6; Length 212;  
Best Local Similarity 60.0%; Pred. No. 15;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NYSPPYNTID 10  
| | | | | | | |  
Db 106 NKSPYHETVD 115

## RESULT 8

US-10-953-349-14759  
; Sequence 14759, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 14759  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-953-349-14759

Query Match 53.6%; Score 37; DB 6; Length 159;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NYSPPYNTIDDL 12  
| | | | | | | |  
Db 5 NRKIYNTMDEI 16

## RESULT 9

US-10-953-349-28574  
; Sequence 28574, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 28574  
; LENGTH: 315  
; TYPE: PRT

; ORGANISM: Triticum aestivum  
US-10-953-349-28574

Query Match 53.6%; Score 37; DB 6; Length 315;  
Best Local Similarity 70.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 YSPYNTID 11  
| | | | | | | |  
Db 263 YSVYNTTRSD 272

## RESULT 10

US-10-953-349-28573  
; Sequence 28573, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 28573  
; LENGTH: 360  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-953-349-28573

Query Match 53.6%; Score 37; DB 6; Length 360;  
Best Local Similarity 70.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 YSPYNTID 11  
| | | | | | | |  
Db 308 YSVYNTTRSD 317

## RESULT 11

US-10-953-349-28572  
; Sequence 28572, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 28572  
; LENGTH: 394  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-953-349-28572

Query Match 53.6%; Score 37; DB 6; Length 394;  
Best Local Similarity 70.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 YSPYNTID 11  
| | | | | | | |  
Db 342 YSVYNTTRSD 351

## RESULT 12

US-10-449-902-31745  
; Sequence 31745, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:

```
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31745
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-31745

Query Match      53.6%; Score 37; DB 6; Length 399;
Best Local Similarity 70.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 YSPYYNTIDD 11
Db      347 YSVYNTSRD 356

RESULT 13
US-10-449-902-55264
; Sequence 55264, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55264
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-55264

Query Match      53.6%; Score 37; DB 6; Length 399;
Best Local Similarity 70.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 YSPYYNTIDD 11
Db      347 YSVYNTSRD 356

RESULT 14
US-10-449-902-55175
; Sequence 55175, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
```

```
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55175
; LENGTH: 1051
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-55175

Query Match      53.6%; Score 37; DB 6; Length 1051;
Best Local Similarity 70.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 SPYYNTIDDL 12
Db      601 TPIQNTIDDL 610

RESULT 15
US-10-985-570-1
; Sequence 1, Application US/10985570
; Publication No. US20060100138A1
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R.
; APPLICANT: Polarek, James W.
; APPLICANT: Yang, Chunlin
; TITLE OF INVENTION: IMPLANTABLE COLLAGEN COMPOSITIONS
; FILE REFERENCE: FP0305 US
; CURRENT APPLICATION NUMBER: US/10/985,570
; CURRENT FILING DATE: 2004-11-10
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-985-570-1

Query Match      53.6%; Score 37; DB 6; Length 1466;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 NYSPYYNTID 10
Db      150 NYSPQYDSYD 159

Search completed: July 5, 2006, 20:53:04
Job time : 7.87379 secs
```



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 27.0291 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-6  
Perfect score: 69  
Sequence: 1 NYSPPYNTIDDL 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H\_COMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE\_COMB.pep.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	100.0	12	US-09-809-003A-6	Sequence 6, Appli
2	53	76.8	420	US-08-961-083-26	Sequence 26, Appl
3	53	76.8	420	US-09-536-784-26	Sequence 26, Appl
4	53	76.8	420	US-09-765-271-26	Sequence 26, Appl
5	53	76.8	420	US-09-765-272A-26	Sequence 26, Appl
6	53	76.8	442	US-09-583-110-3180	Sequence 3180, Ap
7	53	76.8	442	US-09-769-787-121	Sequence 121, App
8	53	76.8	458	US-09-107-433-3143	Sequence 3143, Ap
9	50	72.5	355	US-10-104-047-2943	Sequence 2943, Ap
10	50	72.5	435	US-09-949-016-8415	Sequence 8415, Ap
11	50	72.5	435	US-09-949-016-8792	Sequence 8792, Ap
12	50	72.5	471	US-09-538-092-837	Sequence 837, App
13	50	72.5	593	US-09-538-092-919	Sequence 919, App
14	49	71.0	151	US-09-248-796A-23537	Sequence 23537, A
15	49	71.0	152	US-09-248-796A-27566	Sequence 27566, A
16	46	66.7	242	US-09-248-796A-20477	Sequence 20477, A
17	44	63.8	31	US-07-934-656A-13	Sequence 13, Appl
18	43	62.3	142	US-09-370-838-213	Sequence 213, App
19	43	62.3	142	US-09-854-133-213	Sequence 213, App
20	43	62.3	247	US-10-104-047-3859	Sequence 3859, Ap
21	43	62.3	400	US-07-730-953-2	Sequence 2, Appli
22	43	62.3	442	US-09-949-016-8625	Sequence 8625, Ap
23	43	62.3	1577	US-08-793-824-2	Sequence 2, Appli
24	42	60.9	432	US-08-705-660-18	Sequence 18, Appl
25	42	60.9	432	US-08-989-045-18	Sequence 18, Appl
26	42	60.9	432	US-09-919-172-9	Sequence 9, Appli

27	42	60.9	432	2	US-09-315-355A-18	Sequence 18, Appl
28	41	59.4	185	2	US-09-949-016-9796	Sequence 9796, Ap
29	41	59.4	323	2	US-09-830-230A-644	Sequence 644, App
30	41	59.4	358	2	US-09-830-230A-643	Sequence 643, App
31	40.5	58.7	420	2	US-09-248-796A-17357	Sequence 17357, A
32	40	58.0	31	1	US-07-934-656A-10	Sequence 10, Appl
33	40	58.0	31	1	US-07-934-656A-14	Sequence 14, Appl
34	40	58.0	255	2	US-09-107-433-4006	Sequence 4006, Ap
35	40	58.0	401	2	US-09-248-796A-15983	Sequence 15983, A
36	40	58.0	433	2	US-09-248-796A-20493	Sequence 20493, A
37	40	58.0	501	2	US-09-583-110-3372	Sequence 3372, Ap
38	39	56.5	113	2	US-09-583-110-4936	Sequence 4936, Ap
39	39	56.5	121	2	US-09-107-433-4803	Sequence 4803, Ap
40	39	56.5	124	7	5514582-38	Patent No. 5514582
41	39	56.5	265	1	US-08-177-109A-57	Sequence 57, Appl
42	39	56.5	265	1	US-08-687-706-57	Sequence 16, Appl
43	39	56.5	314	3	US-10-128-713A-16	Sequence 16, Appl
44	39	56.5	364	2	US-09-328-352-6641	Sequence 6641, Ap
45	39	56.5	520	2	US-09-292-858B-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-09-809-003A-6  
; Sequence 6, Application US/09809003A  
; Patent No. 6673351  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-09-809-003A-6

Query Match 100.0%; Score 69; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 6.2e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
|||  
DB 1 NYSPPYNTIDDL 12

RESULT 2  
US-08-961-083-26  
; Sequence 26, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-26

Query Match 76.8%; Score 53; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YSPYNTID 10
DB 368 YSPYNTID 376

RESULT 3
US-09-536-784-26
; Sequence 26, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,271
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/536,784
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:

US-09-536-784-26
Query Match 76.8%; Score 53; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YSPYNTID 10
DB 368 YSPYNTID 376

RESULT 5
US-09-765-272A-26
; Sequence 26, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
```

;; Vaccines  
;;  
;; NUMBER OF SEQUENCES: 454  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue  
;; CITY: Rockville  
;; STATE: Maryland  
;; COUNTRY: USA  
;; ZIP: 20850  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
;; COMPUTER: Dell Latitude C610  
;; OPERATING SYSTEM: Windows 2000  
;; SOFTWARE: ASCII Text  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/765,272A  
;; FILING DATE: 22-Jan-2001  
;; CLASSIFICATION: <Unknown>  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/961,083  
;; FILING DATE: OCT-30-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Lin J. Hymel  
;; REGISTRATION NUMBER: 45,414  
;; REFERENCE/DOCKET NUMBER: PB340P2C2  
;;  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (301) 610-5790  
;; TELEFAX: (301) 309-8439  
;;  
;; INFORMATION FOR SEQ ID NO: 26:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 420 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;;  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 26:  
US-09-765-272A-26  
  
Query Match 76.8%; Score 53; DB 2; Length 420;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 YSPYNTID 10  
Db 368 YSPYNTID 376  
  
RESULT 6  
US-09-583-110-3180  
; Sequence 3180, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 3180  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-3180  
  
Query Match 76.8%; Score 53; DB 2; Length 442;  
Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 YSPYNTID 10  
Db 390 YSPYNTID 398  
  
RESULT 7  
US-09-769-787-121  
; Sequence 121, Application US/09769787  
; Patent No. 6936252  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT APPLICATION NUMBER: US/09/769,787  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 121  
; TYPE: PRT  
; LENGTH: 442  
; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-121  
  
Query Match 76.8%; Score 53; DB 2; Length 442;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 YSPYNTID 10  
Db 390 YSPYNTID 398  
  
RESULT 8  
US-09-107-433-3143  
; Sequence 3143, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489

```
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...458
; SEQUENCE DESCRIPTION: SEQ ID NO: 3143:
US-09-107-433-3143

Query Match 76.8%; Score 53; DB 2; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy 2 YSPYNTID 10
Db 406 YSPYNTID 414

RESULT 9
US-10-104-047-2943
; Sequence 2943, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2943
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2943

Query Match 72.5%; Score 50; DB 2; Length 355;
Best Local Similarity 66.7%; Pred. No. 2.8; Indels 1; Gaps 0;
Matches 8; Conservative 3; Mismatches 1;

Qy 1 NYSPTYNTIDDL 12
Db 139 DYSPYKTIIEEL 150

RESULT 10
US-09-949-016-8415
; Sequence 8415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 837
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Homo sapiens

; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...458
; SEQUENCE DESCRIPTION: SEQ ID NO: 3143:
US-09-107-433-3143

Query Match 76.8%; Score 53; DB 2; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy 2 YSPYNTID 10
Db 406 YSPYNTID 414

RESULT 9
US-10-104-047-2943
; Sequence 2943, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2943
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2943

Query Match 72.5%; Score 50; DB 2; Length 355;
Best Local Similarity 66.7%; Pred. No. 2.8; Indels 1; Gaps 0;
Matches 8; Conservative 3; Mismatches 1;

Qy 1 NYSPTYNTIDDL 12
Db 139 DYSPYKTIIEEL 150

RESULT 10
US-09-949-016-8415
; Sequence 8415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 837
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Homo sapiens

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8415
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8415

Query Match 72.5%; Score 50; DB 2; Length 435;
Best Local Similarity 66.7%; Pred. No. 3.4; Indels 1; Gaps 0;
Matches 8; Conservative 3; Mismatches 1;

Qy 1 NYSPTYNTIDDL 12
Db 166 DYSPYKTIIEEL 177

RESULT 11
US-09-949-016-8792
; Sequence 8792, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8792
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8792

Query Match 72.5%; Score 50; DB 2; Length 435;
Best Local Similarity 66.7%; Pred. No. 3.4; Indels 1; Gaps 0;
Matches 8; Conservative 3; Mismatches 1;

Qy 1 NYSPTYNTIDDL 12
Db 166 DYSPYKTIIEEL 177

RESULT 12
US-09-538-092-837
; Sequence 837, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 837
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Homo sapiens
```

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (0)...(0)  
OTHER INFORMATION: Polypeptide Accession Number P02533  
US-09-538-092-837

Query Match 72.5%; Score 50; DB 2; Length 471;  
Best Local Similarity 66.7%; Pred. No. 3.8;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
:||||:||||  
Db 160 DYSPYKTIIDL 171

## RESULT 13

US-09-538-092-919  
; Sequence 919, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Glot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqformat Version 0.9  
; SEQ ID NO 919  
; LENGTH: 593  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P13645  
US-09-538-092-919

Query Match 72.5%; Score 50; DB 2; Length 593;  
Best Local Similarity 75.0%; Pred. No. 4.8;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
:||||:||||  
Db 195 DYSPYKTIIDL 206

## RESULT 14

US-09-248-796A-23537  
; Sequence 23537, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 23537  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-23537

Query Match 71.0%; Score 49; DB 2; Length 151;  
Best Local Similarity 66.7%; Pred. No. 1.6;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
:||||:||||  
Db 113 NYSYINSLDEL 124

## RESULT 15

US-09-248-796A-27566  
; Sequence 27566, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 27566  
; LENGTH: 152  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-27566

Query Match 71.0%; Score 49; DB 2; Length 152;  
Best Local Similarity 66.7%; Pred. No. 1.6;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NYSPPYNTIDDL 12  
:||||:||||  
Db 117 NYSYINSLDEL 128

Search completed: July 5, 2006, 19:24:27  
Job time : 27.0291 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 12.1359 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928A-7

Perfect score: 47

Sequence: 1 AEAEELYQSK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.80.\*

2: PIR1.\*

3: PIR2.\*

4: PIR3.\*

5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	643	1 KRHU2	keratin 1, type II
2	44	93.6	553	2 I59009	epidermal keratin
3	41	87.2	390	2 A60093	cytokeatin, type I
4	41	87.2	534	2 I37942	keratin 4, type II
5	41	87.2	629	2 A29666	keratin, 65K type
6	40	85.1	581	1 KRMS2	keratin, type II C
7	39	83.0	483	2 A34720	keratin 8, type II
8	38	80.9	384	2 I61769	keratin 6d, type I
9	38	80.9	564	1 KRHUBA	keratin 6a, type I
10	38	80.9	564	1 KRHUBB	keratin 6b, type I
11	38	80.9	564	2 I61770	keratin 6e, type I
12	38	80.9	564	2 I61768	keratin 6c, type I
13	38	80.9	564	2 I61771	keratin 6f, type I
14	38	80.9	638	2 I53169	cytokeatin 2 - hu
15	37	78.7	257	2 I38025	keratin-like prote
16	37	78.7	346	2 S34165	keratin, type II -
17	37	78.7	480	2 A56694	keratin, type II
18	37	78.7	491	2 S05408	keratin, type II,
19	37	78.7	503	2 S29094	keratin, type II,
20	37	78.7	508	1 KRSHU2	keratin type II, m
21	37	78.7	513	2 S08381	hemolysin I secret
22	37	78.7	707	2 D43599	keratin, 58K type
23	36	76.6	310	2 S43855	cytokeatin 8, typ
24	36	76.6	370	2 A25004	keratin, 53K type
25	36	76.6	487	2 J70407	keratin 8, type II
26	36	76.6	489	2 S05474	keratin 8, type II
27	36	76.6	490	2 J50658	cytokeatin EndoA
28	35	74.5	458	2 S09228	intermediate filam
29	35	74.5	461	2 A43782	keratin, type II -

ALIGNMENTS

RESULT 1

KRHU2

keratin 1, type II, cytoskeletal - human

N;Alternate names: 67K type II epidermal keratin; cytokeratin 1

C;Species: Homo sapiens (man)

C;Date: 04-Dec-1986 #sequence\_revision 22-Oct-1999 #text\_change 10-Dec-1999

C;Accession: A22940; A02950; A43342

R;Johnson, L.D.; Idler, W.W.; Zhou, X.M.; Roop, D.R.; Steinert, P.M.

Proc. Natl. Acad. Sci. U.S.A. 82, 1896-1900, 1985

A;Reference number: A22940; MUID:85166239; PMID:2580302

A;Accession: A22940

A;Molecule type: DNA

A;Residues: 1-643 <JOH>

A;Cross-references: UNIPARC:UPI0000173D59; GB:M98776; GB:M11845; GB:M11846; N

R;Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.

J. Biol. Chem. 260, 7142-7149, 1985

A;Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000

late filament subunits.

A;Reference number: A92535; MUID:85207740; PMID:2581964

A;Accession: A02950

A;Molecule type: mRNA

A;Residues: 151-183, 'K', 185-199, 'N', 201-204, 'K', 206-236, 'S', 238-239, 'R', 241-356, 'Y', 358-

, 'S', 638-643 <STE>

A;Cross-references: UNIPARC:UPI000016ABD0; GB:M10938; NID:g186787; PIDN:AAA36153.1; PID:

A;Experimental source: tissue neonatal foreskin

A;Note: the authors translated the codon CUG for residue 476 as Met

R;Chipev, C.C.; Korge, B.P.; Markova, N.; Bale, S.J.; DiGiovanna, J.J.; Compton, J.G.; S

Cell 70, 821-828, 1992

A;Title: A leucine---proline mutation in the H1 subdomain of keratin 1 causes epidermol

A;Reference number: A43342; MUID:92386601; PMID:11381288

A;Accession: A43342

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 144-146, 'P', 148-159, 'P', 161-183, 'K', 185-186 <CHI>

A;Cross-references: UNIPARC:UPI0000173D59; GB:M98776; GB:M11215; GB:M11845; GB:M11846; N

A;Note: sequence extracted from NCBI backbone (NCBIP:112784)

C;Comment: The cytoskeletal and microfibrillar keratins are classified into two types, t

atin IF protein subunit appears to be a heterotetramer of two type I and two type II pro

C;Comment: Keratin 1 is expressed in terminally differentiating epidermis.

C;Genetics:

A;Gene: GDB:KRT1

A;Cross-references: GDB:128198; OMIM:139350

A;Map position: 12q11-12q13

A;Note: defects in this gene may result in epidermolytic hyperkeratosis

C;Complex: heterotetramer of two type I, usually keratin 10 (see PIR:KRHU0), and two typ

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil; heterotetramer; intermediate filament

F;4-143/Region: El and V1 subdomains

F;14-179/Region: H1 subdomain

F;180-492/Domain: rod <ROD>

keratin K7, type I  
keratin, 55K type  
keratin K3 - rabbi  
keratin, 67K type  
hypothetical prote  
keratin type II, h  
RNA-directed RNA p  
keratin 5, type II  
hypothetical prote  
RNA-directed RNA p  
UDPglucose-glycopr  
desmin - African c  
desmin - chicken  
hypothetical prote  
OmpA-related prote  
hypothetical prote

30 35 74.5 469 2 S05602  
31 35 74.5 489 2 B24177  
32 35 74.5 629 2 S42829  
33 35 74.5 645 2 A44861  
34 35 74.5 939 2 AF2503  
35 34 72.3 479 2 A61368  
36 34 72.3 525 1 RRXS15  
37 34 72.3 590 2 A29904  
38 34 72.3 857 2 T05352  
39 34 72.3 878 1 RRXS1B  
40 34 72.3 1447 2 S63669  
41 33 70.2 458 2 A43554  
42 33 70.2 463 1 DMCH  
43 33 70.2 646 2 T34532  
44 33 70.2 1055 2 A87364  
45 32 68.1 249 2 B86231

F:180-214/Region: coil 1A  
F:215-226/Region: linker 1  
F:227-327/Region: coil 1B  
F:328-344/Region: linker 12  
F:345-363/Region: coil 2A  
F:364-371/Region: linker 2  
F:372-492/Region: coil 2B  
F:430/Region: stutter  
F:493-643/Domain: tail <END>  
F:493-643/Domain: H2 subdomain  
F:513-643/Region: V2 and E2 subdomains

Query Match 100.0%; Score 47; DB 1; Length 643;  
Best Local Similarity 100.0%; Pred. No. 0.09; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLYQSK 10  
Db 366 AEASLYQSK 375  
|||||

RESULT 2  
I59009  
epidermal keratin subunit II - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: I59009; A61205  
R:Steinert, P.M.; Parry, D.A.D.; Racoosin, E.L.; Idler, W.W.; Steven, A.C.; Trus, B.L.;  
Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713, 1984  
A:Title: The complete cDNA and deduced amino acid sequence of a type II mouse epidermal  
A:Reference number: I59009; MUID:85014838; PMID:6207530  
A:Accession: I59009  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-553 <RES>  
A:Cross-references: UNIPROT:P50446; UNIPARC:UPI000016CE87; GB:K02108; NID:g198634; PIDN:  
R:Finch, J.; Andrews, K.; Krieg, P.; Fuerstenberger, G.; Slaga, T.; Ootsuyama, A.; Tanoc  
Carcinogenesis 12, 1519-1522, 1991  
A:Title: Identification of a cloned sequence activated during multi-stage carcinogenesis  
A:Reference number: A61205; MUID:91316763; PMID:1713553  
A:Accession: A61205  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 528-553 <FTN>  
A:Cross-references: UNIPARC:UPI00001774C8  
C:Genetics:  
A:Gene: KER2  
C:Superfamily: cytoskeletal keratin

Query Match 93.6%; Score 44; DB 2; Length 553;  
Best Local Similarity 90.0%; Pred. No. 0.33; Indels 0; Gaps 0;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLYQSK 10  
Db 339 AEASLYQTK 348  
|||||

RESULT 3  
A60093  
cytokeratin, type II, early ectodermal - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 30-Sep-1993  
C:Accession: A60093  
R:Charlebois, T.S.; Spencer, D.H.; Tarkington, S.K.; Henry, J.J.; Grainger, R.M.  
Development 108, 33-45, 1990  
A:Title: Isolation of a chick cytokeratin cDNA clone indicative of regional specialization  
A:Reference number: A60093; MUID:90276238; PMID:1693557  
A:Accession: A60093  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-390 <CHA>  
A:Cross-references: UNIPARC:UPI00001774C6

C:Superfamily: cytoskeletal keratin  
C:Keywords: intermediate filament

Query Match 87.2%; Score 41; DB 2; Length 390;  
Best Local Similarity 90.0%; Pred. No. 0.98; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEASLYQSK 10  
Db 198 AEASLYQSK 207  
|||||

## RESULT 4

I37942  
keratin 4, type II, cytoskeletal - human  
N:Alternate names: basic cytokeratin; cytokeratin 4  
C:Species: Homo sapiens (man)  
C:Date: 12-Aug-1996 #sequence\_revision 17-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: I37942; S01068; S21884; A30186; S31662  
R:Wanner, R.; Forster, H.H.; Tilmans, I.; Mischke, D.  
J. Invest. Dermatol. 100, 735-741, 1993  
A:Title: Allelic variations of human keratins K4 and K5 provide polymorphic markers with

A:Reference number: I37942; MUID:93267125; PMID:7684424  
A:Accession: I37942  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A:Residues: 1-127, I', 129-144 <WAN1>  
A:Cross-references: UNIPROT:P19013; UNIPARC:UPI000016AB8C; EMBL:X67683; NID:g34020; PIDN:  
R:Leube, R.E.; Bader, B.L.; Bosch, F.X.; Zimbelmann, R.; Achtsaetter, T.; Franke, W.W.

J. Cell Biol. 106, 1249-1261, 1988

A:Title: Molecular characterization and expression of the stratification-related cytokerat

A:Reference number: S01068; MUID:88198369; PMID:2452170

A:Accession: S01068

A:Molecule type: mRNA

A:Residues: 'S', 128-534 <LEU>

A:Cross-references: UNIPARC:UPI000016ABA3; EMBL:X07695; NID:g34072; PIDN:CAA30534.1; PID:

R:Wanner, R.; Tilmans, I.; Mischke, D.

submitted to the EMBL Data Library, July 1991

A:Reference number: S21884

A:Accession: S21884

A:Molecule type: DNA

A:Residues: 411-534 <WAN2>

A:Cross-references: UNIPARC:UPI000016ABA5; EMBL:X61028; NID:g34076; PIDN:CAA43362.1; PID:

A:Experimental source: allele K4a

C:Genetics:

A:Gene: GDB:KRT4; CYK4

A:Cross-references: GDB:120697; OMIM:123940

A:Map position: 12p11.2-12q11

A:Introns: 463/2; 475/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

F:1-152/Domain: head #status predicted <HEA>

F:153-464/Domain: helical rod #status predicted <ROD>

F:465-534/Domain: tail #status predicted <TAI>

Query Match 87.2%; Score 41; DB 2; Length 534;  
Best Local Similarity 80.0%; Pred. No. 1.4; Indels 0; Gaps 0;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLYQSK 10  
Db 338 AEASLYQTK 347  
|||||

## RESULT 5

A29666  
keratin, 65K type II cytoskeletal - human

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 20-Apr-2001

C:Accession: A29666

R:Klinge, E.M.; Sylvestre, Y.R.; Freedberg, I.M.; Blumenberg, M.

J. Mol. Evol. 24, 319-329, 1987

A:Title: Evolution of keratin genes: different protein domains evolve by different pathways



A;Reference number: A296666; MUID:87254239; PMID:2439698  
A;Accession: A296666  
A;Molecule type: DNA  
A;Residues: 1-629 <KUL>  
A;Cross-references: UNIPARC:UPI0000161C35; GB:X05418; NID:G34040; PIDN:CAA28991.1; PID:G34047  
92; PIDN:CAA28994.1; PID:G1335193; PIDN:CAA28995.1; PID:G1335194; GB:X05421; NID:G34047  
A;Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 2  
A;Note: the complete translation is not annotated in GenBank entries HSKER65A, HSKER65B,  
s mistranslated as a Met initiator codon  
C;Genetics:  
A;Introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 87.2%; Score 41; DB 2; Length 629;  
Best Local Similarity 80.0%; Pred. No. 1.6;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEELYQSK 10  
|||||:|  
Db 387 AEAEELYQTK 396  
|||||:|

RESULT 6  
KRM52  
keratin, type II cytoskeletal - mouse (fragment)  
N;Alternate names: 67-kDa type II keratin  
C;Species: Mus musculus (house mouse)  
C;Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004  
C;Accession: A02951  
R;Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.  
J. Biol. Chem. 260, 7142-7149, 1985  
A;Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000  
late filament subunits.  
A;Reference number: A92535; MUID:85207740; PMID:2581964  
A;Accession: A02951  
A;Molecule type: mRNA  
A;Residues: 1-581 <STE>  
A;Cross-references: UNIPROT:P04104; UNIPARC:UPI0000173D5B; GB:M10937; NID:G198622; PID:G198622  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; intermediate filament  
F;1-106/Domain: head (fragment) <HED>  
F;1-70/Region: V1 subdomain (fragment)  
F;71-106/Region: H1 subdomain  
F;107-419/Domain: rod <ROD>  
F;107-141/Region: coil 1A  
F;142-153/Region: linker 1  
F;154-254/Region: coil 1B  
F;255-271/Region: linker 12  
F;272-290/Region: coil 2A  
F;291-298/Region: linker 2  
F;299-419/Region: coil 2B  
F;357/Region: stutter  
F;420-581/Domain: tail <END>  
F;420-439/Region: H2 subdomain  
F;440-581/Region: V2 and E2 subdomains

Query Match 85.1%; Score 40; DB 1; Length 581;  
Best Local Similarity 80.0%; Pred. No. 2.4;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEAEELYQSK 10  
|||||:|  
Db 293 AEAEELYQSK 302  
|||||:|

RESULT 7  
keratin 8, type II cytoskeletal - human  
C;Species: Homo sapiens (man)  
C;Date: 06-Jul-1990 #sequence\_revision 06-Jul-1990 #text\_change 09-Jul-2004  
C;Accession: A34720; S06888; S12479; I37982; I37983; JS0487; A31942  
R;Yamamoto, R.; Kao, L.C.; McKnight, C.E.; Strauss III, J.F.

A;Reference number: A296666; MUID:87254239; PMID:2439698  
A;Accession: A296666  
A;Molecule type: DNA  
A;Residues: 1-629 <KUL>  
A;Cross-references: UNIPARC:UPI0000161C35; GB:X05418; NID:G34040; PIDN:CAA28991.1; PID:G34047  
92; PIDN:CAA28994.1; PID:G1335193; PIDN:CAA28995.1; PID:G1335194; GB:X05421; NID:G34047  
A;Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 2  
A;Note: the complete translation is not annotated in GenBank entries HSKER65A, HSKER65B,  
s mistranslated as a Met initiator codon  
C;Genetics:  
A;Introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 87.2%; Score 41; DB 2; Length 629;  
Best Local Similarity 80.0%; Pred. No. 1.6;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEELYQSK 10  
|||||:|  
Db 387 AEAEELYQTK 396  
|||||:|

RESULT 6  
KRM52  
keratin, type II cytoskeletal - mouse (fragment)  
N;Alternate names: 67-kDa type II keratin  
C;Species: Mus musculus (house mouse)  
C;Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004  
C;Accession: A02951  
R;Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.  
J. Biol. Chem. 260, 7142-7149, 1985  
A;Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000  
late filament subunits.  
A;Reference number: A92535; MUID:85207740; PMID:2581964  
A;Accession: A02951  
A;Molecule type: mRNA  
A;Residues: 1-581 <STE>  
A;Cross-references: UNIPROT:P04104; UNIPARC:UPI0000173D5B; GB:M10937; NID:G198622; PID:G198622  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; intermediate filament  
F;1-106/Domain: head (fragment) <HED>  
F;1-70/Region: V1 subdomain (fragment)  
F;71-106/Region: H1 subdomain  
F;107-419/Domain: rod <ROD>  
F;107-141/Region: coil 1A  
F;142-153/Region: linker 1  
F;154-254/Region: coil 1B  
F;255-271/Region: linker 12  
F;272-290/Region: coil 2A  
F;291-298/Region: linker 2  
F;299-419/Region: coil 2B  
F;357/Region: stutter  
F;420-581/Domain: tail <END>  
F;420-439/Region: H2 subdomain  
F;440-581/Region: V2 and E2 subdomains

Query Match 85.1%; Score 40; DB 1; Length 581;  
Best Local Similarity 80.0%; Pred. No. 2.4;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEAEELYQSK 10  
|||||:|  
Db 293 AEAEELYQSK 302  
|||||:|

RESULT 7  
keratin 8, type II cytoskeletal - human  
C;Species: Homo sapiens (man)  
C;Date: 06-Jul-1990 #sequence\_revision 06-Jul-1990 #text\_change 09-Jul-2004  
C;Accession: A34720; S06888; S12479; I37982; I37983; JS0487; A31942  
R;Yamamoto, R.; Kao, L.C.; McKnight, C.E.; Strauss III, J.F.

Mol. Endocrinol. 4, 370-374, 1990  
A;Title: Cloning and sequence of cDNA for human placental cytokeratin 8. Regulation of t  
A;Reference number: A34720; MUID:90258929; PMID:1692965  
A;Accession: A34720  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-483 <YAM>  
A;Cross-references: UNIPROT:P05787; UNIPARC:UPI00001774C3; GB:M34225  
R;Leube, R.E.; Bosch, F.X.; Romano, V.; Zimbelmann, R.; Hoefler, H.; Franke, W.W.  
Differentiation 33, 69-85, 1986  
A;Title: Cytokeratin expression in simple epithelia.  
A;Reference number: S06888; MUID:87134779; PMID:2434381  
A;Accession: S06888  
A;Molecule type: mRNA  
A;Residues: 205-483 <LEU>  
A;Cross-references: UNIPARC:UPI000016ABAI; EMBL:X12882  
R;Franke, W.W.  
submitted to the EMBL Data Library, September 1988  
A;Reference number: S12479  
A;Accession: S12479  
A;Molecule type: mRNA  
A;Residues: 205-309, 'I', 311-483 <FRA>  
A;Cross-references: UNIPARC:UPI000016A769; EMBL:X12882; NID:G30312; PIDN:CAA31376.1; PID  
R;Waseem, A.; Alexander, C.M.; Steel, J.B.; Lane, E.B.  
New Biol. 2, 464-478, 1990  
A;Title: Embryonic simple epithelial keratins 8 and 18: chromosomal location emphasizes  
A;Reference number: I37982; MUID:91145351; PMID:1705144  
A;Accession: I37982  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-76, 'V', 78-416, 'S', 418-428, 'G', 430-431, 'S', 433-483 <RES>  
A;Cross-references: UNIPARC:UPI0000161B3E; EMBL:X74929; NID:G400415; PIDN:CAAS2882.1; PI  
A;Accession: I37983  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 151-201, 'VD', 202-271 <RE2>  
A;Cross-references: UNIPARC:UPI000016ABCF; EMBL:X74981; NID:G400417; PIDN:CAAS2916.1; PI  
R;Krauss, S.; Franke, W.W.  
Gene 86, 241-249, 1990  
A;Title: Organization and sequence of the human gene encoding cytokeratin 8.  
A;Reference number: JS0487; MUID:90215304; PMID:1691124  
A;Accession: JS0487  
A;Molecule type: DNA  
A;Residues: 1-76, 'V', 78-428, 'G', 'SOA', 434-483 <KRA>  
A;Cross-references: UNIPARC:UPI000016A82B; GB:M34482; NID:G181572; PIDN:AAA35763.1; PID:  
R;Kulesh, D.A.; Cecena, G.; Darmon, Y.M.; Vasseur, M.; Oshima, R.G.  
Mol. Cell. Biol. 9, 1553-1565, 1989  
A;Title: Posttranslational regulation of keratins: degradation of mouse and human kerati  
A;Reference number: A31942; MUID:89261783; PMID:2471085  
A;Accession: A31942  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-76, 'V', 78-231 <KUL>  
A;Cross-references: UNIPARC:UPI000016A354; GB:M26512; NID:G177796; PIDN:AAA51542.1; PID:  
C;Comment: This protein is one of the type II cytokeratins.  
C;Genetics:  
A;Gene: GDB:KRT8  
A;Cross-references: GDB:I18830; OMIM:148060  
A;Map position: 12p13.2-12q24.1  
A;Introns: 108/3; 178/2; 198/3; 230/3; 327/3; 401/2; 421/1  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; intermediate filament

Query Match 83.0%; Score 39; DB 2; Length 483;  
Best Local Similarity 80.0%; Pred. No. 3.3;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEAEELYQSK 10  
|||||:|  
Db 276 AEAEEMYQIK 285  
|||||:|

RESULT 8

```
I61769
keratin 6d, type II - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: I61769
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.
J. Biol. Chem. 270, 18581-18592, 1995
A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly
A:Reference number: A57398; MUID:95355491; PMID:7543104
A:Accession: I61769
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-384 <RES>
A:Cross-references: UNIPROT:P48667; UNIPARC:UPI000012DB17; GB:L42602; NID:G914823; GB:L4
28; GB:L42608; NID:G914829; GB:L42609; NID:G914830; GB:L42610; NID:G1488252; PIDN:AAB606
C:Genetics:
A:Gene: KRT6D
A:Introns: 72/2; 92/3; 124/3; 179/3; 221/3; 295/2; 307/1
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil

Query Match      80.9%; Score 38; DB 2; Length 384;
Best Local Similarity 80.0%; Pred. No. 4.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      ||||| ||:|
DB      170 AEAESWYQTK 179

RESULT 9
KRHUEA
keratin 6a, type II - human
N:Alternate names: 56-kDa type II keratin; keratin cytoskeletal
C:Species: Homo sapiens (man)
C>Date: 15-Nov-1984 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: A57398; A02944
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.
J. Biol. Chem. 270, 18591-18592, 1995
A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly
A:Reference number: A57398; MUID:95355491; PMID:7543104
A:Accession: A57398
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-564 <TK>
A:Cross-references: UNIPROT:P02538; UNIPARC:UPI000013CD4C; GB:L42575; NID:G908769; GB:L4
74; GB:L42581; NID:G908775; GB:L42582; NID:G908776; GB:L42583; NID:G908777; PIDN:AAC4176
R:Hanukoglu, I.; Fuchs, E.
Cell 33, 915-924, 1983
A:Title: The cDNA sequence of a type II cytoskeletal keratin reveals constant and variab
A:Reference number: A02944; MUID:83259278; PMID:6191871
A:Accession: A02944
A:Molecule type: mRNA
A:Residues: 208-394, 'S', 396-564 <HAN>
A:Cross-references: UNIPARC:UPI000016AB87; GB:J00269; NID:G34068; PIDN:CRA24760.1; PID:G
atin IF protein subunit appears to be a heterotetramer of two type I and two type II pr
C:Genetics:
A:Gene: GDB:KRT6A
A:Cross-references: GDB:128111; OMIM:148041
A:Map position: 12q12-12q21
A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:2-163/Domain: head <HED>
F:127-163/Region: E1 and V1 subdomains
F:162-476/Domain: rod <ROD>
F:199-210/Region: linker 1
F:211-311/Region: coil 1B
F:312-328/Region: linker 12
F:329-347/Region: coil 2A
F:348-355/Region: linker 2

QY      1 AEAESLYQSK 10
      ||||| ||:|
DB      350 AEAESWYQTK 359

Query Match      80.9%; Score 38; DB 1; Length 564;
Best Local Similarity 80.0%; Pred. No. 6.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      ||||| ||:|
DB      350 AEAESWYQTK 359
```

```
F:356-476/Region: coil 2B
F:414/Region: stutler
F:477-564/Domain: tail <END>
F:477-496/Region: H2 subdomain
F:497-564/Region: V2 and E2 subdomains

Query Match      80.9%; Score 38; DB 1; Length 564;
Best Local Similarity 80.0%; Pred. No. 6.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      ||||| ||:|
DB      350 AEAESWYQTK 359

RESULT 10
KRHUEB
keratin 6b, type II - human
N:Alternate names: 56K type II keratin; keratin, cytoskeletal
C:Species: Homo sapiens (man)
C>Date: 04-Dec-1986 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: I61767; A02945
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.
J. Biol. Chem. 270, 18581-18592, 1995
A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly
A:Reference number: A57398; MUID:95355491; PMID:7543104
A:Accession: I61767
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-564 <RES>
A:Cross-references: UNIPROT:P04259; UNIPARC:UPI0000167B38; GB:L42584; NID:G908780; GB:L4
85; GB:L42590; NID:G908786; GB:L42591; NID:G908787; GB:L42592; NID:G908788; PIDN:AAC4176
R:Tyner, A.L.; Eichman, M.J.; Fuchs, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 4683-4687, 1985
A:Title: The sequence of a type II keratin gene expressed in human skin: conservation of
A:Reference number: A02945; MUID:85270392; PMID:2410904
A:Accession: A02945
A:Molecule type: DNA
A:Residues: 2-88, 'AG', '91-115, 'PA', '118, 'LC', '122-158, 'IG', '161-254, 'V', '256-564 <TYN>
A:Cross-references: UNIPARC:UPI0000173D58; GB:M11229; GB:L00205; NID:G186714; PIDN:AAAS94
A:Note: Initiator Met not shown
C:Comment: There are two types of cytoskeletal and microfibrillar keratin, I (acidic) and
bligate heteropolymers, composed of complexes formed by the aggregation of at least one I
C:Genetics:
A:Gene: GDB:KRT6B
A:Cross-references: GDB:128113; OMIM:148042
A:Map position: 12pter-12qter
A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:2-163/Domain: head <HED>
F:127-163/Region: E1 and V1 subdomains
F:128-163/Region: H1 subdomain
F:164-476/Domain: rod <ROD>
F:164-198/Region: coil 1A
F:199-210/Region: linker 1
F:211-311/Region: coil 1B
F:312-328/Region: linker 12
F:329-347/Region: coil 2A
F:348-355/Region: linker 2
F:356-476/Region: coil 2B
F:414/Region: stutler
F:477-564/Domain: tail <END>
F:477-496/Region: H2 subdomain
F:497-564/Region: V2 and E2 subdomains

Query Match      80.9%; Score 38; DB 1; Length 564;
Best Local Similarity 80.0%; Pred. No. 6.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      ||||| ||:|
DB      350 AEAESWYQTK 359
```

## RESULT 11

I61770  
keratin 6e, type II - human  
C:Species: Homo sapiens (man)  
C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
C:Accession: I61770  
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A>Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A:Reference number: A57398; MUID:95355491; PMID:7543104  
A:Accession: I61770  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-564 <RES>  
A:Cross-references: UNIPROT:P48668; UNIPARC:UPI0000167E31; GB:L42611; NID:g908802; PIDN:  
C:Genetics:  
A:Gene: KRT6E  
A>Note: this may not be a distinct gene  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 80.9%; Score 38; DB 2; Length 564;  
Best Local Similarity 80.0%; Pred. No. 6.3;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||  
Db 350 AEASWYQTK 359

## RESULT 12

I61768  
keratin 6c, type II - human  
C:Species: Homo sapiens (man)  
C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
C:Accession: I61768  
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A>Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A:Reference number: A57398; MUID:95355491; PMID:7543104  
A:Accession: I61768  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-564 <RES>  
A:Cross-references: UNIPROT:P48666; UNIPARC:UPI0000167E00; GB:L42593; NID:g908791; GB:L4  
96; GB:L42599; NID:g908797; GB:L42600; NID:g908798; GB:L42601; NID:g908799; PIDN:AAC4176  
C:Genetics:  
A:Gene: KRT6C  
A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 80.9%; Score 38; DB 2; Length 564;  
Best Local Similarity 80.0%; Pred. No. 6.3;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||  
Db 350 AEASWYQTK 359

## RESULT 13

I61771  
keratin 6f, type II - human  
C:Species: Homo sapiens (man)  
C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
C:Accession: I61771  
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A>Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A:Reference number: A57398; MUID:95355491; PMID:7543104

A:Accession: I61771  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-564 <RES>  
A:Cross-references: UNIPROT:P48669; UNIPARC:UPI000013CD50; GB:L42612; NID:g908804; PIDN:  
C:Genetics:  
A:Gene: KRT6F  
A>Note: this may not be a distinct gene  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 80.9%; Score 38; DB 2; Length 564;  
Best Local Similarity 80.0%; Pred. No. 6.3;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||  
Db 350 AEASWYQTK 359

## RESULT 14

I53169  
cyto keratin 2 - human  
C:Species: Homo sapiens (man)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: I53169  
R:Collin, C.; Ohayoun, J.  
Differentiation 51, 137-148, 1992  
A>Title: Suprabasal marker proteins distinguishing keratinizing squamous epithelia: Cyto  
A:Reference number: I53169; MUID:93114504; PMID:1282112  
A:Accession: I53169  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-638 <RES>  
A:Cross-references: UNIPROT:Q01546; UNIPARC:UPI000012DB04; GB:M99063; NID:g181389; PIDN:  
C:Superfamily: cytoskeletal keratin

Query Match 80.9%; Score 38; DB 2; Length 638;  
Best Local Similarity 70.0%; Pred. No. 7.1;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
:|||||  
Db 370 SEAEALYQTK 379

## RESULT 15

I38025  
keratin-like protein - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 09-Jul-2004  
C:Accession: I38025; S60680  
R:Tomasetto, C.; Regnier, C.H.; Moog-Lutz, C.; Mattei, M.G.; Chenard, M.P.; Lidereau, R.  
Genomics 28, 367-376, 1995  
A>Title: Identification of four novel human genes amplified and overexpressed in breast  
A:Reference number: I37080; MUID:96039245; PMID:7490069  
A:Accession: I38025  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-257 <RES>  
A:Cross-references: UNIPROT:Q14533; UNIPARC:UPI000016AD67; EMBL:X80197; NID:g951271; PID

Query Match 78.7%; Score 37; DB 2; Length 257;  
Best Local Similarity 80.0%; Pred. No. 4.4;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||  
Db 43 AEASWYRSK 52

Search completed: July 5, 2006, 19:20:30  
Job time : 13.1359 secs

---

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 102.33 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-7

Perfect score: 47

Sequence: 1 AEAEELYQSK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	643	1 K2C1_HUMAN	P04264 homo sapien
2	41	87.2	147	2 Q4R3D5_MACFA	Q4R3D5 macaca fasc
3	41	87.2	324	2 Q80VP7_MOUSE	Q80VP7 mus musculus
4	41	87.2	381	2 Q32MB2_HUMAN	Q32MB2 homo sapien
5	41	87.2	398	2 Q8NA87_HUMAN	Q8NA87 homo sapien
6	41	87.2	410	2 Q5HYM7_HUMAN	Q5HYM7 homo sapien
7	41	87.2	492	1 K2CO_CHICK	Q93532 gallus gall
8	41	87.2	510	2 Q8WVZ0_HUMAN	Q8WVZ0 homo sapien
9	41	87.2	511	2 Q8WVY9_HUMAN	Q8WVY9 homo sapien
10	41	87.2	520	2 Q8NIN4_HUMAN	Q8NIN4 homo sapien
11	41	87.2	521	2 Q7RTT2_HUMAN	Q7RTT2 homo sapien
12	41	87.2	523	2 Q3SY84_HUMAN	Q3SY84 homo sapien
13	41	87.2	523	2 Q3SV85_HUMAN	Q3SV85 homo sapien
14	41	87.2	523	2 Q96DU2_HUMAN	Q96DU2 homo sapien
15	41	87.2	523	2 Q5RCF8_PONPY	Q3RCY8 pongo pygma
16	41	87.2	524	1 K2C6G_MOUSE	Q9R0H5 mus musculus
17	41	87.2	529	2 Q86V45_HUMAN	Q86V45 homo sapien
18	41	87.2	534	1 K2C4_HUMAN	P19013 homo sapien
19	41	87.2	540	2 Q86V46_HUMAN	Q86V46 homo sapien
20	41	87.2	543	2 Q7RS7_HUMAN	Q7RS7 homo sapien
21	41	87.2	581	2 Q6IFZ5_RAT	Q6IFZ5 rattus norv
22	41	87.2	594	2 Q3UV17_MOUSE	Q3UV17 mus musculus
23	41	87.2	619	2 Q6E1Y9_CANFA	Q6E1Y9 canis famil
24	41	87.2	629	1 K2C3_HUMAN	P12035 homo sapien
25	41	87.2	653	2 Q95LG9_CANFA	Q95LG9 canis famil
26	40	85.1	580	2 Q98UJ3_XENLA	Q98UJ3 xenopus lae
27	40	85.1	625	2 Q6IMF3_RAT	Q6IMF3 rattus norv
28	40	85.1	627	1 K2C1_MOUSE	P04104 mus musculus
29	40	85.1	637	2 Q9D2K8_MOUSE	Q9D2K8 mus musculus
30	39	83.0	279	2 Q7L4M3_HUMAN	Q7L4M3 homo sapien
31	39	83.0	365	2 Q96910_HUMAN	Q96910 homo sapien

#### RESULT 1

ID	K2C1_HUMAN	STANDARD;	PRET;	643 AA.
AC	P04264; Q14720; Q6GSJ0; Q9H298;			
DT	20-MAR-1987, integrated into UniProtKB/Swiss-Prot.			
DT	01-FEB-1996, sequence version 4.			
DT	07-FEB-2006, entry version 76.			
DE	Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1)			
DE	(K1) (67 kDa cyokeratin) (Hair alpha protein).			
GN	Name=KRT1; Synonyms=KRTA;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	NCLEOTIDE SEQUENCE [GENOMIC DNA].			
RP	MEDLINE=85166239; PubMed=2580302;			
RA	Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;			
RT	"Structure of a gene for the human epidermal 67-kDa keratin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).			
RN	[2]			
RP	NCLEOTIDE SEQUENCE [GENOMIC DNA].			
RP	Whitlock N.V., Eady R.A.J., McGrath J.A.;			
RT	"Genomic organization of the human keratin 1 gene.";			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	NCLEOTIDE SEQUENCE [GENOMIC DNA].			
RA	Hatell S.J., Eady R.A.J., Wennerstrand L., Dopping-Hepenstal P.J.,			
RA	Leigh I.M., Munro C., Kelsell D.P.;			
RT	"Novel splice site mutation in keratin 1 underlies mild epidermolytic			
RT	palmoplantar keratoderma in three kindreds.";			
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	NCLEOTIDE SEQUENCE [LARGE SCALE MRNA], AND VARIANT TYR-357.			
RP	MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Hilton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,			
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;			

P05787 homo sapien  
Q53G70 homo sapien  
Q6DHW5 homo sapien  
Q6GMV0 homo sapien  
Q5RB34 pongo pygma  
Q6IFZ9 mus musculus  
Q9H552 homo sapien  
Q5P4C7 homo sapien  
Q5Q2H4 proctoterus  
Q919P5 rana catesb  
Q6E1Z1 canis famil  
Q9CV72 mus musculus  
Q99MH7 mouse  
P48667 homo sapien

RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP PRELIMINARY NUCLEOTIDE SEQUENCE [MRNA] OF 151-643.  
 RX MEDLINE=85207740; PubMed=2581964;  
 RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,  
 RA Roop D.R.;  
 RT "Amino acid sequences of mouse and human epidermal type II keratins of  
 RT Mr 67,000 provide a systematic basis for the structural and functional  
 RT diversity of the end domains of keratin intermediate filament  
 RT subunits.";  
 RL J. Biol. Chem. 260:7142-7149(1985).  
 RN [6]  
 RP PROTEIN SEQUENCE OF 1-7.  
 RC TISSUE=Platelet;  
 RX MEDLINE=22608298; PubMed=12665801; DOI=10.1038/nbt810;  
 RA Gevaert K., Goethals M., Martens L., Van Damme J., Staes A.,  
 RA Thomas G.R., Vandekerckhove J.;  
 RT "Exploring proteomes and analyzing protein processing by mass  
 RT spectrometric identification of sorted N-terminal peptides.";  
 RL Nat. Biotechnol. 21:566-569(2003).  
 RN [7]  
 RP CITRULLINATION.  
 RX MEDLINE=96374388; PubMed=8780679; DOI=10.1006/bbrc.1996.1240;  
 RA Senshu T., Kan S., Ogawa H., Marabe M., Asaga H.;  
 RT "Preferential deimination of keratin K1 and filaggrin during the  
 RT terminal differentiation of human epidermis.";  
 RL Biochem. Biophys. Res. Commun. 225:712-719(1996).  
 RN [8]  
 RP VARIANT EHK PRO-160.  
 RX MEDLINE=92386601; PubMed=1381288; DOI=10.1016/0092-8674(92)90315-4;  
 RA Chipev C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,  
 RA Compton J.G., Steinert P.M.;  
 RT "A leucine-->proline mutation in the H1 subdomain of keratin 1 causes  
 RT epidermolytic hyperkeratosis.";  
 RL Cell 70:821-828(1992).  
 RN [9]  
 RP VARIANT ALLELE 1B 559-GLY--TVR-565 DEL.  
 RX MEDLINE=93107743; PubMed=1281859; DOI=10.1111/1523-1747.ep12614149;  
 RA Korge B.P., Compton J.G., Steinert P.M., Mischke D.;  
 RT "The two size alleles of human keratin 1 are due to a deletion in the  
 RT glycine-rich carboxyl-terminal V2 subdomain.";  
 RL J. Invest. Dermatol. 99:697-702(1992).  
 RN [10]  
 RP VARIANT EHK GLN-489.  
 RX MEDLINE=92376531; PubMed=1380725;  
 RA Rothnagel J.A., Dominy A.M., Dempsey L.D., Longley M.A.,  
 RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;  
 RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic  
 RT hyperkeratosis.";  
 RL Science 257:1128-1130(1992).  
 RN [11]  
 RP VARIANT EHK CYS-481.  
 RX MEDLINE=94216497; PubMed=7512983;  
 RA Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;  
 RT "Genetic mutations in the K1 and K10 genes of patients with  
 RT epidermolytic hyperkeratosis. Correlation between location and disease  
 RT severity.";  
 RL J. Clin. Invest. 93:1533-1542(1994).  
 RN [12]  
 RP VARIANTS EHK GLY-154; SER-187 AND PRO-192.  
 RX MEDLINE=94117869; PubMed=7507151; DOI=10.1111/1523-1747.ep12371725;  
 RA Yang J.-W., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,  
 RA Steinert P.M., Compton J.G.;  
 RT "Mutations in the H1 and IA domains in the keratin 1 gene in  
 RT epidermolytic hyperkeratosis.";  
 RL J. Invest. Dermatol. 102:17-23(1994).  
 RN [13]  
 RP VARIANTS EHK PRO-185 AND SER-187.  
 RX MEDLINE=94117870; PubMed=7507152; DOI=10.1111/1523-1747.ep12371726;  
 RA McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,  
 RA Leigh I.M., Navearia H.A., Higgins C., Harper J.I., Paige D.G.,

RA Morley S.M.;  
 RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous  
 RT congenital ichthyosiform erythroderma (BCIE).";  
 RL J. Invest. Dermatol. 102:24-30(1994).  
 RN [14]  
 RP VARIANT NEPK ILE-73.  
 RX MEDLINE=95096501; PubMed=7528239; DOI=10.1111/1523-1747.ep12412771;  
 RA Kimonis V., Digiovanna J.J., Yang J.-M., Doyle S.Z., Bale S.J.,  
 RA Compton J.G.;  
 RT "A mutation in the VI end domain of keratin 1 in non-epidermolytic  
 RT palmar-plantar keratoderma.";  
 RL J. Invest. Dermatol. 103:764-769(1994).  
 RN [15]  
 RP VARIANT EHK VAL-339.  
 RX MEDLINE=99072666; PubMed=9856846;  
 RX DOI=10.1046/j.1523-1747.1998.00389.x;  
 RA Kremer H., Lavrijsen A.P., McLean W.H.I., Lane E.B., Melchers D.,  
 RA Ruiter D.J., Mariman E.C., Steijlen P.M.;  
 RT "An atypical form of bullous congenital ichthyosiform erythroderma is  
 RT caused by a mutation in the L12 linker region of keratin 1.";  
 RL J. Invest. Dermatol. 111:1224-1226(1998).  
 RN [16]  
 RP VARIANTS AEI PHE-478 AND THR-478.  
 RX MEDLINE=99162195; PubMed=10053007;  
 RA Sybert V.P., Francis J.S., Corden L.D., Smith L.T., Weaver M.,  
 RA Stephens K., McLean W.H.I.;  
 RT "Cyclic ichthyosis with epidermolytic hyperkeratosis: a phenotype  
 RT conferred by mutations in the 2B domain of keratin K1.";  
 RL Am. J. Hum. Genet. 64:732-738(1999).  
 RN [17]  
 RP VARIANT EHK THR-187.  
 RX MEDLINE=99247363; PubMed=10232403;  
 RA Arin M.J., Longley M.A., Kuster W., Huber M., Hohl D., Rothnagel J.A.,  
 RA Roop D.R.;  
 RT "An asparagine to threonine substitution in the 1A domain of keratin  
 RT 1: a novel mutation that causes epidermolytic hyperkeratosis.";  
 RL Exp. Dermatol. 8:124-127(1999).  
 RN [18]  
 RP VARIANT AEI PHE-478.  
 RX MEDLINE=20062276; PubMed=10597140;  
 RA Michael E.J., Schneiderman P., Grossman M.E., Cristiano A.M.;  
 RT "Epidermolytic hyperkeratosis with polycyclic psoriasiform plaques  
 RT resulting from a mutation in the keratin 1 gene.";  
 RL Exp. Dermatol. 8:501-503(1999).  
 RN [19]  
 RP VARIANT EHK PRO-213.  
 RX MEDLINE=20305470; PubMed=10844506;  
 RX DOI=10.1046/j.1365-2230.2000.00625.x;  
 RA Cserhalni-Friedman P.B., Squeo R., Gordon D., Garzon M.,  
 RA Schneiderman P., Grossman M.E., Cristiano A.M.;  
 RT "Epidermolytic hyperkeratosis in a Hispanic family resulting from a  
 RT mutation in the keratin 1 gene.";  
 RL Clin. Exp. Dermatol. 25:241-243(2000).  
 RN [20]  
 RP VARIANT EHK THR-478.  
 RX MEDLINE=20151078; PubMed=10688370;  
 RX DOI=10.1034/j.1600-0625.2000.009001016.x;  
 RA Arin M.J., Longley M.A., Epstein E.H. Jr., Rothnagel J.A., Roop D.R.;  
 RT "Identification of a novel mutation in keratin 1 in a family with  
 RT epidermolytic hyperkeratosis.";  
 RL Exp. Dermatol. 9:16-19(2000).  
 RN [21]  
 RP VARIANT EHK ASP-154.  
 RX MEDLINE=21423058; PubMed=11531804;  
 RX DOI=10.1046/j.1365-2333.2000.04327.x;  
 RA Whittock N.V., Ashton G.H.S., Griffiths W.A.D., Eady R.A.J.,  
 RA McGrath J.A.;  
 RT "New mutations in keratin 1 that cause bullous congenital  
 RT ichthyosiform erythroderma and keratin 2e that cause ichthyosis  
 RT bullosa of Siemens.";  
 RL Br. J. Dermatol. 145:330-335(2001).  
 RN [22]  
 RP INVOLVEMENT IN IHCW.

```
RX MEDLINE=21184521; PubMed=11286616;
RX DOI=10.1046/j.1523-1747.2001.01292.x;

Query Match      100.0%; Score 47; DB 1; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEESLYQSK 10
Db 366 AEAEESLYQSK 375

RESULT 2
Q4R3D5_MACFA
ID Q4R3D5_MACFA PRELIMINARY; PRT; 147 AA.
AC Q4R3D5;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Testis cDNA clone: Q4R3D5-17726, similar to human keratin protein K6irs
DE (K6IRS2)..
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini;
OC Cercopitheidae; Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15944441; DOI=10.1093/molbev/msi187;
RA Osada N., Hirata M., Tanuma R., Kusuda J., Hida M., Suzuki Y.,
RA Sugano S., Gojobori T., Shen C.-K.J., Wu C.I., Hashimoto K.;
RT "Substitution Rate and Structural Divergence of 5'UTR Evolution:
RT Comparative Analysis Between Human and Cynomolgus Monkey cDNAs.";
RL Mol. Biol. Evol. 22:1976-1982(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG International consortium for macaque cDNA sequencing and analysis;
RT "DNA sequences of macaque genes expressed in brain or testis and its
RT evolutionary implications.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB179331; BAE02382.1; -; mRNA.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
SQ SEQUENCE 147 AA; 16859 MW; 9680F74B1FE850F2 CRC64;

Query Match      87.2%; Score 41; DB 2; Length 147;
Best Local Similarity 80.0%; Pred. No. 3.2;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEESLYQSK 10
Db 109 AEAEESLYQSK 118

RESULT 3
Q80VP7_MOUSE
ID Q80VP7_MOUSE PRELIMINARY; PRT; 324 AA.
AC Q80VP7;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE Hypothetical protein MGC54654.
GN Name=MGC54654;
OS Mus musculus (Mouse).

RX MEDLINE=21184521; PubMed=11286616;
RX DOI=10.1046/j.1523-1747.2001.01292.x;

Query Match      100.0%; Score 47; DB 1; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEESLYQSK 10
Db 366 AEAEESLYQSK 375

RESULT 2
Q4R3D5_MACFA
ID Q4R3D5_MACFA PRELIMINARY; PRT; 147 AA.
AC Q4R3D5;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Testis cDNA clone: Q4R3D5-17726, similar to human keratin protein K6irs
DE (K6IRS2)..
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini;
OC Cercopitheidae; Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15944441; DOI=10.1093/molbev/msi187;
RA Osada N., Hirata M., Tanuma R., Kusuda J., Hida M., Suzuki Y.,
RA Sugano S., Gojobori T., Shen C.-K.J., Wu C.I., Hashimoto K.;
RT "Substitution Rate and Structural Divergence of 5'UTR Evolution:
RT Comparative Analysis Between Human and Cynomolgus Monkey cDNAs.";
RL Mol. Biol. Evol. 22:1976-1982(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG International consortium for macaque cDNA sequencing and analysis;
RT "DNA sequences of macaque genes expressed in brain or testis and its
RT evolutionary implications.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB179331; BAE02382.1; -; mRNA.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
SQ SEQUENCE 147 AA; 16859 MW; 9680F74B1FE850F2 CRC64;

Query Match      87.2%; Score 41; DB 2; Length 147;
Best Local Similarity 80.0%; Pred. No. 3.2;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEESLYQSK 10
Db 109 AEAEESLYQSK 118

RESULT 3
Q80VP7_MOUSE
ID Q80VP7_MOUSE PRELIMINARY; PRT; 324 AA.
AC Q80VP7;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE Hypothetical protein MGC54654.
GN Name=MGC54654;
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Eye;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Eye;
RG NIH MGC Project;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC046626; AAH46626.1; -; mRNA.
DR HSSP; P08670; IGK.
DR Ensemble; ENSMUSG00000046162; Mus musculus.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Hypothetical protein.
SQ SEQUENCE 324 AA; 35165 MW; 0E4D7E25144435DD CRC64;

Query Match      87.2%; Score 41; DB 2; Length 324;
Best Local Similarity 80.0%; Pred. No. 7.7;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEESLYQSK 10
Db 300 AEAEESLYQSK 309

RESULT 4
Q32MB2_HUMAN
ID Q32MB2_HUMAN PRELIMINARY; PRT; 381 AA.
AC Q32MB2;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE K6IRS3 protein.
GN Name=K6IRS3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
```

RP NUCLEOTIDE SEQUENCE  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Raha S.S., Loquellano N.A., Toohyuk S., Carninci P., Prange C.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullahy S.J.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Rutterfield Y.S.N., Krzywinski M.I., Skaleka U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=PCR rescued clones;  
 RG NIH MGC Project;  
 RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 DR EMBL; BC109213; AA109214.1; -; mRNA.  
 DR EMBL; BC109212; AA109213.1; -; mRNA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 SQ SEQUENCE 381 AA; 42010 MW; 9049439791F7B1A8 CRC64;  
 Query Match 87.2%; Score 41; DB 2; Length 381;  
 Best Local Similarity 80.0%; Pred. No. 9.2;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLYQSK 10  
 DB 319 AEAELYQTK 328  
 RESULT 5  
 Q8NA87 HUMAN PRELIMINARY; PRT; 398 AA.  
 ID Q8NA87\_HUMAN PRELIMINARY; PRT; 398 AA.  
 AC Q8NA87  
 DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.  
 DT 01-OCT-2002, sequence version 1.  
 DT CDNA FLJ35741 fis, clone TEST12004163, moderately similar to Mus  
 DE musculus type II cytochrome.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Testis;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,

RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,  
 RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotuta T., Kusano J.,  
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,  
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
 RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshihara Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
 RA Moriya S., Moniyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamagaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro A.M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45 (2004).  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 DR EMBL; AK093060; BAC04039.1; -; mRNA.  
 DR HSSP; P08670; 1GK7.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR18993:SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament.  
 SQ SEQUENCE 398 AA; 44698 MW; D5DD16028B8EBA3D CRC64;  
 Query Match 87.2%; Score 41; DB 2; Length 398;  
 Best Local Similarity 80.0%; Pred. No. 9.6;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAESLYQSK 10  
 DB 199 AEAETLYQTK 208  
 RESULT 6  
 Q5HYM7 HUMAN PRELIMINARY; PRT; 410 AA.  
 ID Q5HYM7\_HUMAN PRELIMINARY; PRT; 410 AA.  
 AC Q5HYM7;  
 DT 15-FEB-2005, integrated into UniProtKB/TrEMBL.  
 DT 15-FEB-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 6.  
 DE Hypothetical protein DKFZp666B15196.  
 GN Name=DKFZp666B15196;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Esophagus tumor;  
 RG The German cDNA Consortium;  
 RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,  
 RA Fobo G., Han M., Wiemann S.;  
 RA Submitted (JAN-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC



```
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; BX647095; CA146001.1; -; mRNA.
DR Ensembl; ENSG00000170423; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Hypothetical protein; Intermediate filament.
SQ SEQUENCE 410 AA; 44990 MW; 116584182471A250 CRC64;

Query Match 87.2%; Score 41; DB 2; Length 410;
Best Local Similarity 80.0%; Pred. No. 9.9; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 2;

Qy 1 AEASLYQSK 10
Db 188 AEALYQTK 197
|||||:|:|

RESULT 7
K2CO CHICK
ID K2CO CHICK STANDARD; PRT; 492 AA.
AC 093532;
DT 19-SEP-2003, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1998, sequence version 1.
DT 07-FEB-2006, entry version 30.
DE Keratin, type II cytoskeletal cochlear (Cytokeratin otokeratin).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND DEVELOPMENTAL
STAGE.
RC STRAIN=White leghorn; TISSUE=Basilar papilla;
RX MEDLINE=98409669; PubMed=9736748; DOI=10.1073/pnas.95.19.11400;
RA Heller S., Sheane C.A., Javed Z., Hudspeth A.J.;
RT "Molecular markers for cell types of the inner ear and candidate genes
for hearing disorders.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:11400-11405 (1998).
CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins (By
similarity).
CC -!- TISSUE SPECIFICITY: Abundant in the cochlea of the inner ear,
where it is found in the cells of the tegmentum vasculosum. Lower
levels also found in the heart and forebrain.
CC -!- DEVELOPMENTAL STAGE: Specifically expressed at the late
developmental stages in the cochlea.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AF072698; AAC64042.1; -; mRNA.
DR HSP; P08670; 1GK7.
DR Ensembl; ENSGALG0000010250; Gallus gallus.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament; Keratin.
FT CHAIN 1 492 Keratin, type II cytoskeletal cochlear.
/FTid=PRO_0000063746.
```

```
FT REGION 1 113 Head.
FT REGION 114 421 Rod.
FT REGION 114 149 Coiled 1A.
FT REGION 150 166 Linker 1.
FT REGION 167 259 Coiled 1B.
FT REGION 260 282 Linker 12.
FT REGION 283 421 Coiled 2.
FT REGION 422 492 Tail.
SQ SEQUENCE 492 AA; 53803 MW; A01A66426B9F3941 CRC64;

Query Match 87.2%; Score 41; DB 1; Length 492;
Best Local Similarity 90.0%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10
Db 299 AEASLYQSK 308
|||||:|:|

RESULT 8
Q8WVZO HUMAN
ID Q8WVZO HUMAN PRELIMINARY; PRT; 510 AA.
AC Q8WVZO;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 15.
DE Keratin protein K6lrs.
GN Name=KRT6;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OC NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Porter R.M., McLean W.H.I.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AY033495; AAK55108.1; -; mRNA.
DR HSP; P08670; 1GK7.
DR Ensembl; ENSG00000170486; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 510 AA; 55749 MW; 93AAAF9D75378EE7 CRC64;

Query Match 87.2%; Score 41; DB 2; Length 510;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10
Db 312 AEATLYQTK 321
|||||:|:|

RESULT 9
Q8WVY9 HUMAN
ID Q8WVY9 HUMAN PRELIMINARY; PRT; 511 AA.
AC Q8WVY9;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Keratin protein K6lrs.
```

```

GN Name=KRT6;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Porter R.M., McLean W.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA McLean W.H.1.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY033496; AAK55109.1; -; Genomic_DNA.
DR HSSP; P08670; 1GK7.
DR Ensemble; ENSG00000170486; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR PANTHER; PTHR1893:SFS; Keratin_II.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 511 AA; 55877 MW; 61E930C382063731 CRC64;

Query Match 87.2%; Score 41; DB 2; Length 511;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEELYQSK 10
Db |||||:|:|:|
312 AEAEELYQSK 321

RESULT 10
Q8IN14 HUMAN PRELIMINARY; PRT; 520 AA.
AC Q8IN14;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-MAR-2006, entry version 20.
DE CDNA FLJ39100 fis. clone NTONG2001587, moderately similar to KERATIN,
DE TYPE II CYTOSKELETAL 4.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=tongue;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Oka T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayaashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Houchi T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

```

```

RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aoteuka S.,
RA Yoshikawa Y., Matsuura H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsuura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK096419; BAC04782.1; -; mRNA.
DR HSSP; P08670; 1GK7.
DR Ensemble; ENSG00000170423; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR011598; HLH_DNA_bd.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin I.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR1893:SFS; Keratin_II.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament.
SQ SEQUENCE 520 AA; 56820 MW; 1B3C01F3F9DC9158 CRC64;

Query Match 87.2%; Score 41; DB 2; Length 520;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEELYQSK 10
Db |||||:|:|:|
298 AEAEELYQSK 307

RESULT 11
Q7RTT2 HUMAN PRELIMINARY; PRT; 521 AA.
AC Q7RTT2;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-MAR-2006, entry version 13.
DE Keratin 5b.
DE Name=K5B;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21539571; PubMed=11683385;
RA Hesse M., Magin T.M., Weber K.;
RT "Genes for intermediate filament proteins and the draft sequence of
RT the human genome: novel keratin genes and a surprisingly high number
RT of pseudogenes related to keratin genes 8 and 18."
RL J. Cell Sci. 114:2569-2575(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.

```

RA Hesse M., Magin T.M., Weber K.;  
RL Submitted (Oct-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- MISCELLANEOUS: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ third party annotation (TPA) entry.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; BK000646; DAA00377.1; -; Genomic\_DNA.  
DR HSP; P08670; 1GK7.  
DR Ensemble; ENSG00000170423; Homo sapiens.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR003054; Keratin\_II.  
DR PANTHER; PTHR1893:SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 521 AA; 56965 MW; 94A0C125C9077F08 CRC64;  
  
Query Match 87.2%; Score 41; DB 2; Length 521;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 AEAESLYQSK 10  
Db 298 AEAALYQTK 307  
||||:||||:  
-----  
RESULT 12  
ID Q3SY84 HUMAN PRELIMINARY; PRT; 523 AA.  
AC Q3SY84;  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 11-OCT-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE Keratin 6 IRS.  
GN Name=KRT6IRS;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RG NIH MGC Project;  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>

RG NIH MGC Project;  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; BC103918; AA103919.1; -; mRNA.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR003054; Keratin\_II.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 523 AA; 57221 MW; 7861868BFF46D16 CRC64;  
  
Query Match 87.2%; Score 41; DB 2; Length 523;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 AEAESLYQSK 10  
Db 317 AEAALYQTK 326  
||||:||||:  
-----  
RESULT 13  
ID Q3SY85 HUMAN PRELIMINARY; PRT; 523 AA.  
AC Q3SY85;  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 11-OCT-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE Keratin 6 IRS.  
GN Name=KRT6IRS;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RG NIH MGC Project;  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>

```
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BG103917; AAI03918.1; -; mRNA.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 523 AA; 57249 MW; 797F5644FF3A73C6 CRC64;

Query Match 87.2%; Score 41; DB 2; Length 523;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEALYQSK 10
Db 317 AEAEALYQTK 326

RESULT 14
Q96DU2 HUMAN PRELIMINARY; PRT; 523 AA.
AC Q96DU2
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 17.
DE Keratin 6 IRS.
GN Names=KRT6IRS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Scalp;
RA Rogers M.A., Langbein L., Winter H., Praetzel S., Ehmann C.,
RA Schweizer J.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ308599; CAC43429.1; -; mRNA.
DR HSSP; P08670; 1GK7.
DR Ensembl; ENSG00000139648; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 523 AA; 57292 MW; 797F5655EE3A62D7 CRC64;

Query Match 87.2%; Score 41; DB 2; Length 523;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEALYQSK 10
Db 317 AEAEALYQTK 326

RESULT 15
Q5RCY8_PONPY
Q5RCY8_PONPY PRELIMINARY; PRT; 523 AA.
```

```
AC Q5RCY8;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DE 07-FEB-2006, entry version 7.
DE Hypothetical protein DKFZp468M2112.
GN Name=DKFZp468M2112;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Heart;
RG The German cDNA Consortium;
RA Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA Newes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR858130; CAH90369.1; -; mRNA.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Hypothetical protein; Intermediate filament.
SQ SEQUENCE 523 AA; 57249 MW; 797F5644FF3A73C6 CRC64;

Query Match 87.2%; Score 41; DB 2; Length 523;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEALYQSK 10
Db 317 AEAEALYQTK 326

Search completed: July 5, 2006, 19:01:32
Job time : 102.43 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 86.8932 Seconds

(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928A-7

Perfect score: 47

Sequence: 1 AEAESLYQSK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

9: Geneseqp2005ss.\*

10: Geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	10	6	AAO26387 Psoriasis
2	47	100.0	10	9	ADV04412 Leishmani
3	47	100.0	12	8	AD119066 Rat pepti
4	47	100.0	644	8	ADQ17549 Human sof
5	47	100.0	644	9	AEA15501 Human pol
6	41	87.2	55	4	ABB42646 Peptide #
7	41	87.2	55	4	AAM36459 Peptide #
8	41	87.2	55	4	ABB25991 Protein #
9	41	87.2	55	4	AAM76351 Human bon
10	41	87.2	55	4	AAM63536 Human bra
11	41	87.2	55	4	ABG58058 Human liv
12	41	87.2	55	5	ABG45643 Human pep
13	41	87.2	314	7	AD136103 Human NOV
14	41	87.2	327	7	ADJ69669 Human hea
15	41	87.2	336	7	ADJ70556 Human hea
16	41	87.2	398	7	ADB65234 Human pro
17	41	87.2	420	8	ADP29879 Human sec
18	41	87.2	508	8	ADS10914 Human the
19	41	87.2	511	5	ABG94648 Human NOV
20	41	87.2	511	7	ADL36000 Human NOV
21	41	87.2	511	9	AEA15498 Human pol
22	41	87.2	520	7	ADB75609 Prostata
23	41	87.2	520	7	ADC31845 Human nov

24	41	87.2	520	7	ADM05027 Human pro
25	41	87.2	520	9	AEC87957 Human GDN
26	41	87.2	521	5	ABG94649 Human NOV
27	41	87.2	521	5	ABB78802 Human NOV
28	41	87.2	521	7	ADL36002 Human NOV
29	41	87.2	523	6	AAE32116 Human cyt
30	41	87.2	523	9	AEA15496 Human NOV
31	41	87.2	523	5	ABG60120 Human DIT
32	41	87.2	526	5	ABJ04655 Protein o
33	41	87.2	529	5	ABB83482 Human cyt
34	41	87.2	529	9	AEA15497 Human pol
35	41	87.2	534	5	ABP43881 57kd kera
36	41	87.2	534	7	ADF09537 Human ker
37	41	87.2	534	9	AEA15503 Human pol
38	41	87.2	536	5	ABJ04646 Protein o
39	41	87.2	540	9	AEA15499 Human pol
40	41	87.2	542	5	ABB78799 Human NOV
41	41	87.2	543	9	AEA20229 Novel hum
42	41	87.2	549	5	ABG94647 Human NOV
43	41	87.2	549	7	ADL35998 Human NOV
44	41	87.2	608	6	ADA54878 Human pro
45	41	87.2	629	9	AEA15502 Human pol

#### ALIGNMENTS

RESULT 1  
AAO26387  
ID AAO26387 standard; peptide; 10 AA.  
XX  
AC AAO26387;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 7.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN W0200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
PA (AKIV-) AKIVA LLC.  
XX  
PI O'daly JA;  
XX  
DR WPI; 2003-018763/01.  
XX  
PT New polypeptide and immunogenic variants comprising amino acid sequences  
PT of particulate antigens, useful for the treatment and clinical remission  
PT of psoriasis.  
XX  
PS Claim 1; Page 43; 56pp; English.  
XX  
CC The invention relates to a polypeptide comprising an isolated amino acid  
CC sequence or immunogenic variants selected from any of 14 fully defined  
CC sequences of 7-16 amino acids, given in the specification. The  
CC immunotherapeutic agents and a pharmaceutical compositions comprising  
CC polynucleotides and vectors of the invention are useful for the treatment  
CC and clinical remission of psoriasis. The isolated nucleic acids are  
CC useful as probes. The sequences of the invention can be used in the  
CC treatment of disorders by gene therapy. This sequence represents one of  
CC the 14 immunotherapeutic peptides of the invention  
XX

```
SQ Sequence 10 AA;
Query Match 100.0%; Score 47; DB 6; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLYQSK 10
   |||||
Db 1 AEASLYQSK 10

RESULT 2
ADV04412
ID ADV04412 standard; peptide; 10 AA.
XX
AC ADV04412;
XX
DT 24-FEB-2005 (first entry)
XX
DE Leishmania immunotherapeutic peptide SEQ ID NO: 7.
XX
KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;
KW psoriasis.
XX
OS Leishmania sp.
XX
PN US2004241168-A1.
XX
PD 02-DEC-2004.
XX
PF 09-FEB-2004; 2004US-00774928.
XX
PR 16-MAR-2001; 2001US-00809003.
PR 17-OCT-2003; 2003US-00687892.
XX
(PDAL/) ODALY J A.
PA
PI Odaly JA;
XX
WPI; 2005-011563/01.
XX
Inhibiting selectively T-cell rolling in human, by administering compound
PT interfering with cutaneous lymphocyte antigen-E selectin interaction and
PT leukocyte function associated antigen-1 interaction.
XX
Claim 18; SEQ ID NO 7; 21pp; English.
XX
The invention relates to a novel method for selectively inhibiting T-cell
rolling in a human host, comprising administering a compound that
selectively interferes with the cutaneous lymphocyte antigen (CLA)-E
selectin interaction and leukocyte function associated antigen (LFA)-
1/intercellular adhesion molecule (ICAM) and very late antigen
(VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound
includes an immunotherapeutic agent, which comprises a purified protein
extract that is isolated by diethylaminoethyl Sephadex chromatography of
a Nonidet P-40 insoluble particulate antigen fraction derived from
isolated killed cells of amastigotes from one or more species of the
Leishmania genus, where the particulate antigen fraction is solubilized
with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied
to diethylaminoethyl Sephadex and is eluted with a solution comprising
0.1 M sodium chloride, 8 M urea and 0.025 M
Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract
includes polypeptides having apparent molecular weights of 73, 80 and 82
kDa, after total reduction and alkylation. The species is L. amazonensis,
L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the
invention has antipsoriatic and immunostimulant activity, and inhibits T-
cell rolling by interfering with CLA-E selectin interaction and LFA-
1/ICAM and VLA/VCAM interactions. The method is useful for selectively
inhibiting T-cell rolling in a human host, and for treating psoriasis.
The present sequence represents a Leishmania peptide used in the
invention.
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 47; DB 6; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLYQSK 10
   |||||
Db 1 AEASLYQSK 10

RESULT 3
ADI19066
ID ADI19066 standard; peptide; 12 AA.
XX
AC ADI19066;
XX
DT 15-APR-2004 (first entry)
XX
DE Rat peptide #1 used in the exemplification of the invention.
XX
KW Detrusor estrogen-regulated protein; DERP; DERP-related disorder;
KW impaired detrusor contractility; urinary retention; Alzheimer's disease;
KW cardiovascular disease; osteoporosis; smooth muscle function; neurotropic;
KW neuroprotective; osteopathic; immunostimulant; therapy; rat.
XX
OS Rattus norvegicus.
XX
PN US2003232374-A1.
XX
PD 18-DEC-2003.
XX
PF 22-MAY-2003; 2003US-00444575.
PR 23-MAY-2002; 2002US-0382830P.
XX
(KUCH/) KUCHEL G A.
PA (ZHUQ/) ZHU Q.
XX
PI Kuchel GA, Zhu Q;
XX
WPI; 2004-052160/05.
XX
Novel purified immunogenic detrusor estrogen-regulated polypeptide
PT fragment useful for generating antibodies and as screening reagents.
XX
Example 11; SEQ ID NO 16; 48pp; English.
XX
The present invention relates to a immunogenic polypeptide fragment
comprising ten to fifty consecutive amino acids of a fully defined
detrusor estrogen-regulated protein (DERP). The invention is useful for
diagnosing DERP-related disorder such as impaired detrusor contractility,
urinary retention, Alzheimer's disease, cardiovascular disease,
osteoporosis, or their combinations. The invention is also useful for
generating antibodies, as reagents in diagnostic assays, as identifiers
of other gene products involved in the regulation of bladder and smooth
muscle function, as screening reagents useful for detecting compounds
that are used in the regulation of bladder and smooth muscle function,
and as pharmaceutical agents useful for treating symptoms relating to
bladder and smooth muscle function. The present sequence is rat peptide
used in the exemplification of the invention.
XX
SQ Sequence 12 AA;
Query Match 100.0%; Score 47; DB 8; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEASLYQSK 10
   |||||
Db 3 AEASLYQSK 12

RESULT 4
Sequence 10 AA;
```

ADQ17549  
 ID ADQ17549 standard; protein; 644 AA.  
 XX AC ADQ17549;  
 XX DT 26-AUG-2004 (first entry)  
 XX DE Human soft tissue sarcoma-upregulated protein - SEQ ID 366.  
 XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
 XX OS Homo sapiens.  
 XX PN W02004048938-A2.  
 XX PD 10-JUN-2004.  
 XX PF 26-NOV-2003; 2003WO-US038193.  
 XX PR 26-NOV-2002; 2002US-0429739P.  
 XX PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX PI Aziz N, Ginsburg WM, Zlotnik A;  
 XX WIPI; 2004-441208/41.  
 XX PT Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 XX PS Example 2; SEQ ID NO 366; 210pp; English.  
 XX CC The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC protein of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 XX SQ Sequence 644 AA;  
 Query Match 100.0%; Score 47; DB 8; Length 644;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAEESLYQSK 10  
 DB 367 AEAEESLYQSK 376  
 RESULT 5  
 AEAL5501  
 ID AEAL5501 standard; protein; 644 AA.  
 XX AC AEAL5501;  
 XX DT 28-JUL-2005 (first entry)  
 XX DE Human polypeptide #120.  
 XX KW Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
 KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
 KW cytostatic; neoplasm.

OS Homo sapiens.  
 XX WO2005047534-A2.  
 XX PD 26-MAY-2005.  
 XX PF 15-OCT-2004; 2004WO-EP011599.  
 XX PR 28-OCT-2003; 2003EP-00024565.  
 XX PA (FARB ) BAYER HEALTHCARE AG.  
 XX PI Wirtz R, Munnes M;  
 XX WIPI; 2005-372393/39.  
 XX DR N-PSDB; AEAL5423.  
 XX PT Predicting a response to cancer treatment by detecting at least 2  
 PT markers, which are genes or genomic nucleic acid sequences that are  
 PT located on one chromosomal region, which is altered in malignant  
 PT neoplasia.  
 XX PS Claim 7; SEQ ID NO 463; 464pp; English.  
 XX CC The invention relates to a method of predicting response to cancer  
 CC treatment comprising detection of at least 2 markers, where the markers  
 CC are genes and fragments or genomic nucleic acid sequences that are  
 CC located on one chromosomal region, which is altered in malignant  
 CC neoplasia. The invention also relates to a method for the prediction,  
 CC diagnosis or prognosis of malignant neoplasia, methods for detecting  
 CC deregulations in malignant neoplasia and breast cancer, a method of  
 CC determining the phenotype of a cell or tissue, a method for identifying  
 CC genomic regions which are altered on the chromosomal level and encode  
 CC genes that are linked by function and are differentially expressed in  
 CC malignant neoplasia and breast cancer, methods of screening for agents  
 CC which regulate the activity of a polypeptide or a polynucleotide and  
 CC antibodies that specifically bind to a full length or partial  
 CC polypeptide. The method is useful for predicting response to cancer  
 CC treatment. The methods and compositions are useful for predicting,  
 CC diagnosing, prognosing, preventing or treating malignant neoplasia  
 CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
 CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
 CC lung cancer. This sequence represents a human polypeptide used in the  
 CC scope of the invention.  
 XX SQ Sequence 644 AA;  
 Query Match 100.0%; Score 47; DB 9; Length 644;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAEESLYQSK 10  
 DB 367 AEAEESLYQSK 376  
 RESULT 6  
 ABB42646  
 ID ABB42646 standard; peptide; 55 AA.  
 XX AC ABB42646;  
 XX DT 04-FEB-2002 (first entry)  
 XX DE Peptide #10152 encoded by human foetal liver single exon probe.  
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
 XX OS Homo sapiens.  
 XX PN W0200157277-A2.  
 XX PD 09-AUG-2001.

```

XX PF 30-JAN-2001; 2001WO-US0000669.
XX PR 04-FEB-2000; 2000US-01803112P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
  gene expression in human fetal liver.
XX Claim 27; SEQ ID NO 35281; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
  human gene expression in a sample derived from human foetal liver. The
  single exon nucleic acid probes may be used for predicting, measuring and
  displaying gene expression in samples derived from human fetal liver. The
  present sequence is a peptide encoded by a single exon nucleic acid probe
  of the invention. Note: The sequence data for this patent did not form
  part of the printed specification, but was obtained in electronic format
  directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 55 AA;

Query Match 87.2%; Score 41; DB 4; Length 55;
Best Local Similarity 80.0%; Pred. No. 2.5;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10
Db 46 AEAALYQTK 55
|||||:|:|

RESULT 7
AAM36459
ID AAM36459 standard; protein; 55 AA.
AC AAM36459;
XX 17-OCT-2001 (first entry)
DT Peptide #10496 encoded by probe for measuring placental gene expression.
DE Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US0000663.
XX 04-FEB-2000; 2000US-01803112P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
  hearts.
XX Claim 15; SEQ ID NO 27761; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
  measuring human gene expression in a sample derived from human heart (see
  ABA21535-ABA411305). The present sequence is a protein encoded by one such
  probe. The probes may be used for predicting, measuring and displaying
  gene expression in samples derived from the human heart via microarrays.

```

---

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
  gene expression in human placenta.
XX Claim 27; SEQ ID NO 36728; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP:
  see AA131315-AA157546). The present sequence is a peptide encoded by one
  such probe. The probes are useful for producing a microarray for
  predicting, measuring and displaying gene expression in samples derived
  from human placenta. The probes are useful for antenatal diagnosis of
  human genetic disorders
XX SQ Sequence 55 AA;

Query Match 87.2%; Score 41; DB 4; Length 55;
Best Local Similarity 80.0%; Pred. No. 2.5;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10
Db 46 AEAALYQTK 55
|||||:|:|

RESULT 8
ABB25991
ID ABB25991 standard; protein; 55 AA.
XX ABB25991;
XX 23-JAN-2002 (first entry)
DT Protein #7990 encoded by probe for measuring heart cell gene expression.
DE Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX Homo sapiens.
XX WO200157274-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US0000666.
XX 04-FEB-2000; 2000US-01803112P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
  hearts.
XX Claim 15; SEQ ID NO 27761; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
  measuring human gene expression in a sample derived from human heart (see
  ABA21535-ABA411305). The present sequence is a protein encoded by one such
  probe. The probes may be used for predicting, measuring and displaying
  gene expression in samples derived from the human heart via microarrays.

```



CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 55 AA;

Query Match 87.2%; Score 41; DB 4; Length 55;  
Best Local Similarity 80.0%; Pred. No. 2.5;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
DB 46 AEAELYQTK 55  
||||:||||:

RESULT 9  
AMW76351  
ID AAMW76351 standard; protein; 55 AA.  
XX  
AC AAMW76351;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 36657.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PD WPI; 2001-488900/53.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PD WPI; 2001-488900/53.  
XX  
PF Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human bone marrow.  
XX  
PS Example 4; SEQ ID NO 36657; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention  
XX  
SQ Sequence 55 AA;

Query Match 87.2%; Score 41; DB 4; Length 55;  
Best Local Similarity 80.0%; Pred. No. 2.5;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10

DB 46 AEAELYQTK 55  
||||:||||:

RESULT 10  
AAM63536  
ID AAM63536 standard; protein; 55 AA.  
XX  
AC AAM63536;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 35641.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000667.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PD WPI; 2001-483446/52.  
XX  
PF Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
PS Example 4; SEQ ID NO 35641; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention  
XX  
SQ Sequence 55 AA;

Query Match 87.2%; Score 41; DB 4; Length 55;  
Best Local Similarity 80.0%; Pred. No. 2.5;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
DB 46 AEAELYQTK 55  
||||:||||:

RESULT 11  
ABG58058  
ID ABG58058 standard; peptide; 55 AA.  
XX  
AC ABG58058;  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Human liver peptide, SEQ ID No 36706.

KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157273-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US000664.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-48898/53.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.  
 PT  
 PS Claim 27; SEQ ID NO 36706; 658pp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 55 AA;  
 Query Match 87.2%; Score 41; DB 4; Length 55;  
 Best Local Similarity 80.0%; Pred. No. 2.5;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AEAEELYQSK 10  
 Db 46 AEAEELYQTK 55  
 RESULT 12  
 ABG45643  
 ID ABG45643 standard; peptide; 55 AA.  
 XX  
 AC ABG45643;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 35308.  
 XX  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US000665.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.  
 PT  
 PS Claim 27; SEQ ID NO 35308; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC ; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberosus sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 55 AA;  
 Query Match 87.2%; Score 41; DB 5; Length 55;  
 Best Local Similarity 80.0%; Pred. No. 2.5;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEALYQSK 10  
 Db 46 AEAEALYQTK 55

RESULT 13  
 ADL36103  
 ID ADL36103 standard; protein; 314 AA.

XX  
 AC ADL36103;  
 DT 20-MAY-2004 (first entry)  
 XX Human NOVX-related polypeptide #29.

DE Human; NOVX; Alzheimer's disease; Parkinson's disease; stroke; epilepsy;  
 KW multiple sclerosis; addiction; anxiety; pain; acne; alopecia;  
 KW inflammation; rheumatoid arthritis; AIDS; cancer; psoriasis;  
 KW hypertension; renal disorder; bone disease; haematopoietic disorder;  
 KW wound; bacterial infection; viral infection; fungal infection;  
 KW protozoal infection; urinary retention; osteoporosis;  
 KW myocardial infarction; diabetes; ulcer; cirrhosis; depression.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2003207800-A1.  
 XX  
 XX 06-NOV-2003.  
 XX  
 XX 13-NOV-2001; 2001US-00015115.  
 XX  
 XX 08-NOV-1999; 99US-0164240P.  
 PR 13-NOV-2000; 2000US-0248153P.  
 PR 17-NOV-2000; 2000US-0249598P.  
 PR 02-FEB-2001; 2001US-0266127P.  
 PR 16-FEB-2001; 2001US-0269562P.  
 PR 10-JUL-2001; 2001US-0304348P.  
 PR 31-JUL-2001; 2001US-0309261P.  
 PR 17-AUG-2001; 2001US-0313283P.  
 XX  
 XX (MALV/) MALYANKAR U M.  
 PA (SHEN/) SHENOY S G.  
 PA (SPYI/) SPYTEK K A.  
 PA (ZERR/) ZERHUSEN B D.  
 PA (PATT/) PATTURAJAN M.  
 PA (GUOX/) GUO X.  
 PA (KEKU/) KEKUDA R.  
 PA (GANG/) GANGOLLI E A.  
 PA (SHIM/) SHIMKETS R A.  
 PA (TAUP/) TAUPIER R J.  
 PA (LILL/) LI L.  
 PA (PADI/) PADIGARU M.  
 XX  
 XX Malyankar UM, Shenoy SG, Spytek KA, Zerhusen BD, Patturajan M;  
 PI Guo X, Kekuda R, Gangolli EA, Shimkets RA, Taupier RJ, Li L;  
 PI Padigaru M;  
 XX  
 XX WPI; 2003-875894/81.  
 XX  
 XX New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOVX-associated disorders (e.g. stroke, epilepsy,  
 PT AIDS, pain, diabetes or cancer) and in chromosome mapping, tissue typing  
 PT or pharmacogenomics.  
 XX  
 XX Disclosure; SEQ ID NO 206; 233pp; English.  
 XX  
 XX The invention relates to human NOVX polypeptides and the polynucleotides  
 CC encoding them. The invention also relates to antibodies that bind  
 CC immunospecifically to the polypeptides. The NOVX polypeptides,  
 CC polynucleotides and antibodies are useful in diagnosing, treating or  
 CC preventing NOVX-associated disorders such as Alzheimer's disease,

CC Parkinson's disease, stroke, epilepsy, multiple sclerosis, addiction,  
 CC anxiety, pain, acne, alopecia, inflammation, rheumatoid arthritis, AIDS,  
 CC cancer, psoriasis, hypertension, renal disorders, bone diseases,  
 CC haematopoietic disorders, wounds, infection (e.g. bacterial, viral,  
 CC fungal or protozoal), urinary retention, osteoporosis, myocardial  
 CC infarction, diabetes, ulcer, cirrhosis or depression. The polypeptides  
 CC are also useful as vaccines. This sequence represents a human NOVX-  
 CC related polypeptide of the invention.

XX  
 SQ Sequence 314 AA;  
 Query Match 87.2%; Score 41; DB 7; Length 314;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEALYQSK 10  
 Db 189 AEAEALYQTK 198

RESULT 14  
 ADJ69669  
 ID ADJ69669 standard; protein; 327 AA.

XX  
 AC ADJ69669;  
 XX  
 XX 06-MAY-2004 (first entry)  
 XX  
 XX Human heat mitochondrial protein as a therapeutic target SeqID1475.  
 XX  
 XX mitochondrial; human; screening assay; diabetes mellitus;  
 KW Huntington's disease; osteoarthritis;  
 KW Leber's hereditary optic neuropathy; LHON;  
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
 KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;  
 KW osteopathic; ophthalmological; cytostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2003087768-A2.  
 PN  
 XX  
 XX 23-OCT-2003.  
 XX  
 XX 04-APR-2003; 2003WO-US010870.  
 XX  
 XX 12-APR-2002; 2002US-0372843P.  
 PR 17-JUN-2002; 2002US-0389987P.  
 PR 20-SEP-2002; 2002US-0412418P.  
 XX  
 XX (MITO-) MITOKOR.  
 PA (BUCK-) BUCK INST AGE RES.  
 XX  
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GW;  
 PI Warnock DE;  
 XX  
 XX WPI; 2003-845369/78.  
 DR  
 XX  
 XX Identifying a mitochondrial target for drug screening assays and for  
 PT treating diseases associated with altered mitochondrial function,  
 PT comprises detecting a modified polypeptide in a sample and correlating  
 PT with the disease.  
 XX  
 XX Claim 1; SEQ ID NO 1475; 180pp; English.  
 XX  
 XX This invention relates to novel mitochondrial targets that can be used  
 CC for therapeutic intervention in treating a disease associated with  
 CC altered mitochondrial function. Specifically, it refers to a method for  
 CC identifying proteins of the human heart mitochondrial proteome that are  
 CC useful for drug screening assays, as well as therapeutic targets. The  
 CC present invention describes a method for identifying such proteins that  
 CC can be used in the treatment of various diseases associated with altered  
 CC mitochondrial function including diabetes mellitus, Huntington's disease,

CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, nootropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytotatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.  
XX  
SQ Sequence 327 AA;

Query Match 87.2%; Score 41; DB 7; Length 327;  
Best Local Similarity 80.0%; Pred. No. 17;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 AEAESLYQSK 10  
| | | | | | | | | |  
Db 131 AEAALYQTK 140

## RESULT 15

ADJ70556  
ID ADJ70556 standard; protein; 336 AA.

XX  
AC ADJ70556;

DT 06-MAY-2004 (first entry)

XX Human heat mitochondrial protein as a therapeutic target SeqID2362.

DE Mitochondrial; human; screening assay; diabetes mellitus;

XX Huntington's disease; osteoarthritis;

KW Leber's hereditary optic neuropathy; LHON;

KW Mitochondrial encephalopathy lactic acidosis and stroke; MELAS;

KW Myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;

KW Neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;

KW Osteopathic; ophthalmological; cytotatic.

XX Homo sapiens.

OS  
XX WO2003087768-A2.

PN  
XX 23-OCT-2003.

XX 04-APR-2003; 2003WO-US010870.

XX 12-APR-2002; 2002US-0372843P.

PR 17-JUN-2002; 2002US-0389987P.

PR 20-SEP-2002; 2002US-0412418P.

XX (MITO-) MITOKOR.

PA (BUCK-) BUCK INST AGE RES.

XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX Warnock DE;

XX WPI; 2003-845369/78.

XX Identifying a mitochondrial target for drug screening assays and for

PT treating diseases associated with altered mitochondrial function,

PT comprises detecting a modified polypeptide in a sample and correlating

XX with the disease.

PS Claim 1; SEQ ID NO 2362; 180pp; English.

XX This invention relates to novel mitochondrial targets that can be used

CC for therapeutic intervention in treating a disease associated with

CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, nootropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytotatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.  
XX  
SQ Sequence 336 AA;

Query Match 87.2%; Score 41; DB 7; Length 336;  
Best Local Similarity 80.0%; Pred. No. 17;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 AEAESLYQSK 10  
| | | | | | | | | |  
Db 235 AEAALYQTK 244

Search completed: July 5, 2006, 18:43:53  
Job time : 87.8932 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 73.2039 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-7  
Perfect score: 47  
Sequence: 1 AEAESLYQSK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main.\*  
1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	10	5	US-10-774-928-7
2	47	100.0	12	4	US-10-444-575-16
3	47	100.0	644	5	US-10-723-860-366
4	41	87.2	55	3	US-09-864-761-41289
5	41	87.2	222	4	US-10-231-913-263
6	41	87.2	315	4	US-10-231-913-259
7	41	87.2	318	4	US-10-231-913-261
8	41	87.2	327	4	US-10-408-765A-1475
9	41	87.2	336	4	US-10-015-115-126
10	41	87.2	336	4	US-10-408-765A-2362
11	41	87.2	398	4	US-10-104-047-3388
12	41	87.2	398	6	US-11-072-512-3388
13	41	87.2	441	4	US-10-015-115-124
14	41	87.2	441	4	US-10-231-913-85
15	41	87.2	441	4	US-10-231-913-115
16	41	87.2	511	4	US-10-015-115-46
17	41	87.2	520	4	US-10-205-823-433
18	41	87.2	520	4	US-10-108-260A-3712
19	41	87.2	520	6	US-11-051-454-433
20	41	87.2	521	3	US-09-976-782-10
21	41	87.2	521	4	US-10-015-115-48
22	41	87.2	523	4	US-10-015-115-123
23	41	87.2	523	4	US-10-231-913-83
24	41	87.2	523	4	US-10-231-913-113
25	41	87.2	523	4	US-10-473-574-14
26	41	87.2	524	4	US-10-015-115-125
27	41	87.2	524	4	US-10-231-913-84

28	41	87.2	524	4	US-10-231-913-114
29	41	87.2	526	4	US-10-231-913-32
30	41	87.2	529	4	US-10-250-613-11
31	41	87.2	529	6	US-11-124-368A-182
32	41	87.2	534	3	US-09-976-782-86
33	41	87.2	534	3	US-09-976-782-87
34	41	87.2	534	4	US-10-231-913-86
35	41	87.2	534	4	US-10-231-913-87
36	41	87.2	534	5	US-10-367-057-38
37	41	87.2	536	4	US-10-231-913-14
38	41	87.2	542	3	US-09-976-782-4
39	41	87.2	549	4	US-10-015-115-44
40	41	87.2	608	4	US-10-094-749-2446
41	40	85.1	390	4	US-10-408-765A-2878
42	39	83.0	279	6	US-11-013-684-21
43	39	83.0	398	5	US-10-883-020-5
44	39	83.0	422	3	US-09-779-307-18
45	39	83.0	429	5	US-10-450-763-49905

ALIGNMENTS

RESULT 1

US-10-774-928-7  
; Sequence 7, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-7

Query Match 100.0%; Score 47; DB 5; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AEAESLYQSK	10
Db	1	AEAESLYQSK	10

RESULT 2

US-10-444-575-16  
; Sequence 16, Application US/10444575  
; Publication No. US20030232374A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Connecticut Health Center  
; APPLICANT: Kuchel, George A  
; APPLICANT: Zhu, Qing  
; TITLE OF INVENTION: Compositions and Methods Relating to Detrusor Estrogen-Regulated  
; TITLE OF INVENTION: Protein (DERP)  
; FILE REFERENCE: UCT-0035  
; CURRENT APPLICATION NUMBER: US/10/444,575  
; CURRENT FILING DATE: 2003-05-22  
; PRIOR APPLICATION NUMBER: US 60/382,830  
; PRIOR FILING DATE: 2002-05-23  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 16  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-10-444-575-16

```
Query Match      100.0%; Score 47; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.031; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

Qy      1 AEAESLYQSK 10
      |||||
Db      3 AEAESLYQSK 12

RESULT 3
US-10-723-860-366
; Sequence 366, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 366
; LENGTH: 644
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-366

Query Match      100.0%; Score 47; DB 5; Length 644;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AEAESLYQSK 10
      |||||
Db      367 AEAESLYQSK 376

RESULT 4
US-09-864-761-41289
; Sequence 41289, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR FILING DATE: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41289
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC024196.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EST_HUMAN HIT: BE747524.1, EVALUE 2.00e-18
; OTHER INFORMATION: SWISSPROT HIT: P19013, EVALUE 2.00e-19
US-09-864-761-41289
```

```
Query Match      87.2%; Score 41; DB 3; Length 55;
Best Local Similarity 80.0%; Pred. No. 2.2;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 AEAESLYQSK 10
      |||||
Db      46 AEAESLYQTK 55
```

```
RESULT 5
US-10-231-913-263
; Sequence 263, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
```

APPLICANT: Spaderna, Steven K.  
APPLICANT: Zerhusen, Bryan D.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-216  
CURRENT APPLICATION NUMBER: US/10/231,913  
CURRENT FILING DATE: 2002-08-30  
PRIOR APPLICATION NUMBER: 60/251,660  
PRIOR FILING DATE: 2000-12-06  
PRIOR APPLICATION NUMBER: 60/255,029  
PRIOR FILING DATE: 2000-12-12  
PRIOR APPLICATION NUMBER: 60/260,326  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: 60/263,800  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: 60/318,712  
PRIOR FILING DATE: 2001-09-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 263  
LENGTH: 222  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-231-913-263

Query Match 87.2%; Score 41; DB 4; Length 222;  
Best Local Similarity 80.0%; Pred. No. 9.6;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
Db 138 AEAEALYQTK 147  
|||||:|||||

RESULT 6  
US-10-231-913-259  
Sequence 259, Application US/10231913  
Publication No. US20040005576A1  
GENERAL INFORMATION:  
APPLICANT: Guo, Xiaojia S.  
APPLICANT: Li, Li  
APPLICANT: Patturajan, Meera  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Casman, Stacie J.  
APPLICANT: Malyankar, Uriel M.  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Vernet, Corine A.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Shenoy, Suresh G.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Edinger, Schlomit  
APPLICANT: Peyman, John A.  
APPLICANT: Stone, David J.  
APPLICANT: Ellerman, Karen  
APPLICANT: Gangolli, Esha A.  
APPLICANT: Boldog, Ference L.  
APPLICANT: Colman, Steven D.  
APPLICANT: Eisen, Andrew J.  
APPLICANT: Liu, Xiaohong  
APPLICANT: Padigar, Muralidhara  
APPLICANT: Spaderna, Steven K.  
APPLICANT: Zerhusen, Bryan D.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-216  
CURRENT APPLICATION NUMBER: US/10/231,913  
CURRENT FILING DATE: 2002-08-30  
PRIOR APPLICATION NUMBER: 60/251,660  
PRIOR FILING DATE: 2000-12-06

PRIOR APPLICATION NUMBER: 60/255,029  
PRIOR FILING DATE: 2000-12-12  
PRIOR APPLICATION NUMBER: 60/260,326  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: 60/263,800  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: 60/269,942  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/286,183  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/313,627  
PRIOR FILING DATE: 2001-08-20  
PRIOR APPLICATION NUMBER: 60/318,712  
PRIOR FILING DATE: 2001-09-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 259  
LENGTH: 315  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-231-913-259

Query Match 87.2%; Score 41; DB 4; Length 315;  
Best Local Similarity 80.0%; Pred. No. 14;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
Db 190 AEAEALYQTK 199  
|||||:|||||

RESULT 7  
US-10-231-913-261  
Sequence 261, Application US/10231913  
Publication No. US20040005576A1  
GENERAL INFORMATION:  
APPLICANT: Guo, Xiaojia S.  
APPLICANT: Li, Li  
APPLICANT: Patturajan, Meera  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Casman, Stacie J.  
APPLICANT: Malyankar, Uriel M.  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Vernet, Corine A.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Shenoy, Suresh G.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Edinger, Schlomit  
APPLICANT: Peyman, John A.  
APPLICANT: Stone, David J.  
APPLICANT: Ellerman, Karen  
APPLICANT: Gangolli, Esha A.  
APPLICANT: Boldog, Ference L.  
APPLICANT: Colman, Steven D.  
APPLICANT: Eisen, Andrew J.  
APPLICANT: Liu, Xiaohong  
APPLICANT: Padigar, Muralidhara  
APPLICANT: Spaderna, Steven K.  
APPLICANT: Zerhusen, Bryan D.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-216  
CURRENT APPLICATION NUMBER: US/10/231,913  
CURRENT FILING DATE: 2002-08-30  
PRIOR APPLICATION NUMBER: 60/251,660  
PRIOR FILING DATE: 2000-12-06  
PRIOR APPLICATION NUMBER: 60/255,029  
PRIOR FILING DATE: 2000-12-12  
PRIOR APPLICATION NUMBER: 60/260,326  
PRIOR FILING DATE: 2001-01-08  
PRIOR APPLICATION NUMBER: 60/263,800  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: 60/269,942  
PRIOR FILING DATE: 2001-02-20

```
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/318,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 261
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-913-261

Query Match      87.2%; Score 41; DB 4; Length 318;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:||||:|
Db      195 AEAEALYQTK 204

RESULT 8
US-10-408-765A-1475
; Sequence 1475, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1475
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1475

Query Match      87.2%; Score 41; DB 4; Length 327;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:||||:|
Db      131 AEAEALYQTK 140

RESULT 9
US-10-015-115-126
; Sequence 126, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesha
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shimkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li

; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-211
; CURRENT APPLICATION NUMBER: US/10/015,115
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-115-126

Query Match      87.2%; Score 41; DB 4; Length 336;
Best Local Similarity 80.0%; Pred. No. 15;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:||||:|
Db      235 AEAEALYQTK 244

RESULT 10
US-10-408-765A-2362
; Sequence 2362, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2362
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2362

Query Match      87.2%; Score 41; DB 4; Length 336;
Best Local Similarity 80.0%; Pred. No. 15;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:||||:|
Db      235 AEAEALYQTK 244
```



RESULT 11  
US-10-04-047-3388  
; Sequence 3388, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: HI-A0105  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3388  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-04-047-3388

Query Match 87.2%; Score 41; DB 4; Length 398;  
Best Local Similarity 80.0%; Pred. No. 18;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLYQSK 10  
||||:||||:  
Db 199 AEATLYQTK 208

RESULT 12  
US-11-072-512-3388  
; Sequence 3388, Application US/11072512  
; Publication No. US20060029945A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KETICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: 084335-0191  
; CURRENT APPLICATION NUMBER: US/11/072,512  
; CURRENT FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/350,978  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: JP 2001-379298  
; PRIOR FILING DATE: 2001-11-05  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3388  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-072-512-3388

Query Match 87.2%; Score 41; DB 6; Length 398;  
Best Local Similarity 80.0%; Pred. No. 18;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLYQSK 10

Db 199 AEATLYQTK 208  
||||:||||:  
RESULT 13  
US-10-015-115-124  
; Sequence 124, Application US/10015115  
; Publication No. US20030207800A1  
; GENERAL INFORMATION:  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Shenoy, Suresh G  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Kekuda, Ramesha  
; APPLICANT: Gangolli, Baha A  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Taupier, Raymond J  
; APPLICANT: Li, Li  
; APPLICANT: Padigaru, Muralidhara  
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
; FILE REFERENCE: 21402-211  
; CURRENT APPLICATION NUMBER: US/10/015,115  
; CURRENT FILING DATE: 2002-09-23  
; PRIOR APPLICATION NUMBER: 60/248,153  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: 60/249,598  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/264,240  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/286,127  
; PRIOR FILING DATE: 2001-02-02  
; PRIOR APPLICATION NUMBER: 60/269,562  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: 60/304,348  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/309,261  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/313,283  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 205  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 124  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-015-115-124

Query Match 87.2%; Score 41; DB 4; Length 441;  
Best Local Similarity 80.0%; Pred. No. 20;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAESLYQSK 10  
||||:||||:  
Db 235 AEATLYQTK 244

RESULT 14  
US-10-231-913-85  
; Sequence 85, Application US/10231913  
; Publication No. US20040005576A1  
; GENERAL INFORMATION:  
; APPLICANT: Guo, Xiaojia S.  
; APPLICANT: Li, Li  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Edinger, Schlomit  
; APPLICANT: Feyman, John A.  
; APPLICANT: Stone, David J.  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Boldog, Ference L.  
; APPLICANT: Colman, Andrew J.  
; APPLICANT: Eisen, Andrew J.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Spaderma, Steven K.  
; APPLICANT: Zerhusen, Bryan D.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-216  
; CURRENT APPLICATION NUMBER: US/10/231,913  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: 60/251,660  
; PRIOR FILING DATE: 2000-12-06  
; PRIOR APPLICATION NUMBER: 60/255,029  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: 60/260,326  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: 60/263,800  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/269,942  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/286,183  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/313,627  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/318,712  
; PRIOR FILING DATE: 2001-09-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 85  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-231-913-85

Query Match 87.2%; Score 41; DB 4; Length 441;  
Best Local Similarity 80.0%; Pred. No. 20;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEALYQSK 10  
DB 235 AEAEALYQTK 244  
|||||:

RESULT 15  
US-10-231-913-115  
; Sequence 115, Application US/10231913  
; Publication No. US20040005576A1  
; GENERAL INFORMATION:  
; APPLICANT: Guo, Xiaojia S.  
; APPLICANT: Li, Li  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Edinger, Schlomit  
; APPLICANT: Feyman, John A.  
; APPLICANT: Stone, David J.  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Boldog, Ference L.

; APPLICANT: Colman, Steven D.  
; APPLICANT: Eisen, Andrew J.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Spaderma, Steven K.  
; APPLICANT: Zerhusen, Bryan D.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-216  
; CURRENT APPLICATION NUMBER: US/10/231,913  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: 60/251,660  
; PRIOR FILING DATE: 2000-12-06  
; PRIOR APPLICATION NUMBER: 60/255,029  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: 60/260,326  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: 60/263,800  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/269,942  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/286,183  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/313,627  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/318,712  
; PRIOR FILING DATE: 2001-09-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 115  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-231-913-115

Query Match 87.2%; Score 41; DB 4; Length 441;  
Best Local Similarity 80.0%; Pred. No. 20;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAEALYQSK 10  
DB 235 AEAEALYQTK 244  
|||||:

Search completed: July 5, 2006, 20:51:59  
Job time: 73.2039 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 5.72816 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-7

Perfect score: 47

Sequence: 1 AEAEELYQSK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /EMC\_Celerra\_SID33/prodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 2: /EMC\_Celerra\_SID33/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /EMC\_Celerra\_SID33/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /EMC\_Celerra\_SID33/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 5: /EMC\_Celerra\_SID33/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 6: /EMC\_Celerra\_SID33/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /EMC\_Celerra\_SID33/prodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /EMC\_Celerra\_SID33/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	41	87.2	520	7	US-11-293-697-3712
2	39	83.0	483	7	US-11-105-233-180
3	35	74.5	469	6	US-10-505-928-457
4	34	72.3	639	6	US-10-449-902-41447
5	33	70.2	221	6	US-10-449-902-30681
6	33	70.2	984	6	US-10-449-902-56454
7	32	68.1	282	6	US-10-449-902-54246
8	32	68.1	576	6	US-10-953-349-32301
9	32	68.1	626	6	US-10-449-902-54896
10	32	68.1	645	7	US-11-354-079-3
11	32	68.1	662	6	US-10-449-902-54986
12	32	68.1	689	6	US-10-953-349-32300
13	32	68.1	689	6	US-10-449-902-35350
14	32	68.1	745	6	US-10-953-349-32239
15	31	66.0	220	6	US-10-449-902-30877
16	31	66.0	220	6	US-10-449-902-33167
17	31	66.0	378	7	US-11-174-307B-5390
18	31	66.0	423	6	US-10-449-902-37702
19	31	66.0	546	6	US-10-471-571A-3376
20	31	66.0	599	7	US-11-293-697-2702
21	31	66.0	637	7	US-11-293-697-2455
22	31	66.0	869	6	US-10-471-571A-4926
23	30	63.8	402	7	US-11-174-307B-282
24	30	63.8	450	7	US-11-293-697-4090
25	30	63.8	474	6	US-10-449-902-31122

26 30 63.8 527 6 US-10-449-902-54681 Sequence 54681, A  
27 30 63.8 754 6 US-10-449-902-41136 Sequence 41136, A  
28 29 61.7 265 6 US-10-449-902-54848 Sequence 54848, A  
29 29 61.7 347 6 US-10-449-902-50919 Sequence 50919, A  
30 29 61.7 403 7 US-11-270-040-2 Sequence 2, Appli  
31 29 61.7 442 6 US-10-449-902-37142 Sequence 37142, A  
32 29 61.7 458 6 US-10-449-902-35627 Sequence 35627, A  
33 29 61.7 466 7 US-11-289-102-289 Sequence 289, App  
34 29 61.7 514 6 US-10-449-902-52901 Sequence 52901, A  
35 29 61.7 514 6 US-10-449-902-53911 Sequence 53911, A  
36 29 61.7 551 7 US-11-270-040-4 Sequence 4, Appli  
37 29 61.7 778 6 US-10-953-349-10588 Sequence 10588, A  
38 29 61.7 792 6 US-10-953-349-10587 Sequence 10587, A  
39 29 61.7 1061 6 US-10-449-902-41065 Sequence 41065, A  
40 29 61.7 1191 6 US-10-449-902-54909 Sequence 54909, A  
41 29 61.7 1511 6 US-10-449-902-43668 Sequence 43668, A  
42 28 59.6 95 6 US-10-953-349-35135 Sequence 35135, A  
43 28 59.6 95 7 US-11-225-400-12 Sequence 12, Appli  
44 28 59.6 100 7 US-11-197-712-468 Sequence 468, App  
45 28 59.6 155 7 US-11-293-697-3730 Sequence 3730, Ap

#### ALIGNMENTS

##### RESULT 1

US-11-293-697-3712  
; Sequence 3712, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cdna  
; FILE REFERENCE: HI-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; PRIOR FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3712  
; LENGTH: 520  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-293-697-3712

Query Match 87.2%; Score 41; DB 7; Length 520;  
Best Local Similarity 80.0%; Pred. No. 0.51;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAEELYQSK 10

Db 298 AEAEELYQTK 307

##### RESULT 2

US-11-105-233-180  
; Sequence 180, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:

; APPLICANT: Thiagalingam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; FILE REFERENCE: 1657/2001  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; CURRENT FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-105-233-180

```
Query Match      83.0%; Score 39; DB 7; Length 483;
Best Local Similarity 80.0%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:|
Db      276 AEAESMYQIK 285

RESULT 3
US-10-505-928-457
; Sequence 457, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 457
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-505-928-457

Query Match      74.5%; Score 35; DB 6; Length 469;
Best Local Similarity 70.0%; Pred. No. 8.5;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:|
Db      277 AEAESMYQTK 286

RESULT 4
US-10-449-902-41447
; Sequence 41447, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41447
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-41447

Query Match      72.3%; Score 34; DB 6; Length 639;
Best Local Similarity 60.0%; Pred. No. 20;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:|
Db      474 AEAEGVIQTR 483

RESULT 5
US-10-449-902-41447
; Sequence 41447, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41447
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-41447

Query Match      70.2%; Score 33; DB 6; Length 984;
Best Local Similarity 60.0%; Pred. No. 52;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:|
Db      182 AHADSLYER 191

RESULT 6
US-10-449-902-56454
; Sequence 56454, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56454
; LENGTH: 984
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-56454

Query Match      70.2%; Score 33; DB 6; Length 984;
Best Local Similarity 60.0%; Pred. No. 52;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
      |||||:|
Db      182 AHADSLYER 191

RESULT 7
US-10-449-902-54246
; Sequence 54246, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
```

; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 54246  
; LENGTH: 282  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-54246

Query Match 68.1%; Score 32; DB 6; Length 282;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAEELYQ 8  
Db 26 AEESLYE 33

RESULT 8  
US-10-953-349-32301  
; Sequence 32301, Application US/10953349  
; Publication No. US20060107348A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 32301  
; LENGTH: 576  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-953-349-32301

Query Match 68.1%; Score 32; DB 6; Length 576;  
Best Local Similarity 70.0%; Pred. No. 46;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AEAEELYQ 10  
Db 378 AEQALYQSK 387

RESULT 9  
US-10-449-902-54896  
; Sequence 54896, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11

; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 54896  
; LENGTH: 626  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-54896

Query Match 68.1%; Score 32; DB 6; Length 626;  
Best Local Similarity 70.0%; Pred. No. 51;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AEAEELYQSK 10  
Db 493 AEQALYQSK 502

RESULT 10  
US-11-354-079-3  
; Sequence 3, Application US/11354079  
; Publication No. US20060141522A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: Heregulin-Like Factor  
; FILE REFERENCE: PF383D2  
; CURRENT APPLICATION NUMBER: US/11/354,079  
; CURRENT FILING DATE: 2006-02-15  
; PRIOR APPLICATION NUMBER: 10/609,370  
; PRIOR FILING DATE: 2003-07-01  
; PRIOR APPLICATION NUMBER: 09/097,681  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/049,942  
; PRIOR FILING DATE: 1997-06-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-354-079-3

Query Match 68.1%; Score 32; DB 7; Length 645;  
Best Local Similarity 66.7%; Pred. No. 53;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EAESLYQSK 10  
Db 239 EAELYQKR 247

RESULT 11  
US-10-449-902-54986  
; Sequence 54986, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 54986  
; LENGTH: 662  
; TYPE: PRT  
; ORGANISM: Oryza sativa

## US-10-449-902-54986

Query Match 68.1%; Score 32; DB 6; Length 662;  
Best Local Similarity 70.0%; Pred. No. 54;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
|||:||||  
Db 481 AEQEALYGSK 490

## RESULT 12

US-10-953-349-32300  
; Sequence 32300, Application US/10953349  
; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 32300  
; LENGTH: 689  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-953-349-32300

Query Match 68.1%; Score 32; DB 6; Length 689;  
Best Local Similarity 70.0%; Pred. No. 57;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
|||:||||  
Db 491 AEQEALYGSK 500

## RESULT 13

US-10-449-902-35350  
; Sequence 35350, Application US/10449902  
; Publication No. US20060123505A1

## GENERAL INFORMATION:

; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 35350  
; LENGTH: 689  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-35350

Query Match 68.1%; Score 32; DB 6; Length 689;  
Best Local Similarity 70.0%; Pred. No. 57;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
|||:||||  
Db 493 AEQEALYGSK 502

## RESULT 14

US-10-953-349-32299  
; Sequence 32299, Application US/10953349  
; Publication No. US20060107345A1

## GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 32299  
; LENGTH: 745  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-10-953-349-32299

Query Match 68.1%; Score 32; DB 6; Length 745;  
Best Local Similarity 70.0%; Pred. No. 62;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10  
|||:||||  
Db 547 AEQEALYGSK 556

## RESULT 15

US-10-449-902-30877  
; Sequence 30877, Application US/10449902  
; Publication No. US20060123505A1

## GENERAL INFORMATION:

; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 30877  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-30877

Query Match 66.0%; Score 31; DB 6; Length 220;  
Best Local Similarity 85.7%; Pred. No. 25;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLY 7  
|||:||||  
Db 149 AEAESMY 155

Search completed: July 5, 2006, 20:53:03  
Job time : 6.72816 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 22.5243 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-7

Perfect score: 47

Sequence: 1 AEAESLYQSK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/iaa/5 COMB.pcp.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/iaa/6 COMB.pcp.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/iaa/7 COMB.pcp.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/iaa/H COMB.pcp.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/iaa/PCUTUS COMB.pcp.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/iaa/RE COMB.pcp.\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	10	2	US-09-809-003A-7
2	47	100.0	637	2	US-09-949-016-8152
3	47	100.0	643	2	US-09-538-092-844
4	41	87.2	398	2	US-10-104-047-3388
5	41	87.2	608	2	US-10-094-749-2446
6	39	83.0	443	2	US-09-949-016-10582
7	39	83.0	482	2	US-09-538-092-858
8	39	83.0	483	2	US-09-919-497-79
9	39	83.0	502	2	US-09-949-016-11033
10	38	80.9	551	1	US-09-067-351-2
11	38	80.9	551	2	US-09-360-490-2
12	38	80.9	564	2	US-09-949-016-6628
13	38	80.9	569	2	US-09-949-016-11035
14	38	80.9	569	2	US-09-949-016-11036
15	37	78.7	235	2	US-09-620-405B-487
16	37	78.7	235	2	US-09-604-287A-487
17	37	78.7	235	2	US-09-834-759-487
18	37	78.7	235	2	US-10-076-622-487
19	37	78.7	235	2	US-10-124-805-487
20	37	78.7	274	2	US-09-949-016-9095
21	37	78.7	493	2	US-09-949-016-6203
22	37	78.7	505	2	US-09-620-405B-478
23	37	78.7	505	2	US-09-620-405B-478
24	37	78.7	505	2	US-09-604-287A-478
25	37	78.7	505	2	US-09-604-287A-485
26	37	78.7	505	2	US-09-834-759-478

#### ALIGNMENTS

##### RESULT 1

US-09-809-003A-7  
; Sequence 7, Application US/09809003A

; Patent No. 6673351

; GENERAL INFORMATION:

; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission

; TITLE OF INVENTION: Psoriasis

; FILE REFERENCE: 302178

; CURRENT APPLICATION NUMBER: US/09/809,003A

; CURRENT FILING DATE: 2001-03-16

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 7

; LENGTH: 10

; TYPE: PRT

; ORGANISM: Leishmania

; US-09-809-003A-7

Query Match 100.0%; Score 47; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.004;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEAESLYQSK 10

Db 1 AEAESLYQSK 10

##### RESULT 2

US-09-949-016-8152

; Sequence 8152, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8152

; LENGTH: 637

Sequence 485, App  
Sequence 478, App  
Sequence 478, App  
Sequence 478, App  
Sequence 478, App  
Sequence 485, App  
Sequence 485, App  
Sequence 478, App  
Sequence 485, App  
Sequence 478, App  
Sequence 478, App  
Sequence 173, App  
Sequence 173, App  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 11034, A  
Sequence 3, Appli  
Sequence 9096, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 41, Appli  
Sequence 7138, Ap

27 37 78.7 505 2 US-09-834-759-485  
28 37 78.7 505 2 US-09-590-751A-478  
29 37 78.7 505 2 US-09-551-621-478  
30 37 78.7 505 2 US-09-551-621A-478  
31 37 78.7 505 2 US-10-076-622-478  
32 37 78.7 505 2 US-10-076-622-485  
33 37 78.7 505 2 US-10-124-805-478  
34 37 78.7 505 2 US-10-124-805-485  
35 37 78.7 528 2 US-09-949-016-9678  
36 37 78.7 687 2 US-09-302-626B-173  
37 37 78.7 707 2 US-08-772-270A-4  
38 37 78.7 707 2 US-09-062-126-4  
39 35 74.5 384 2 US-09-949-016-11034  
40 35 74.5 459 2 US-09-077-606-3  
41 35 74.5 476 2 US-09-949-016-9096  
42 35 74.5 546 1 US-09-067-351-1  
43 35 74.5 546 2 US-09-360-490-1  
44 35 74.5 645 2 US-09-919-172-41  
45 34 72.3 136 2 US-09-543-681A-7138

```
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8152

Query Match      100.0%; Score 47; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
DB      367 AEAESLYQSK 376
|||||

RESULT 3
US-09-538-092-844
; Sequence 844, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 844
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P04264
US-09-538-092-844

Query Match      100.0%; Score 47; DB 2; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
DB      366 AEAESLYQSK 375
|||||

RESULT 4
US-10-104-047-3388
; Sequence 3388, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3388
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3388

Query Match      87.2%; Score 41; DB 2; Length 398;
Best Local Similarity 80.0%; Pred. No. 3.9;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AEAESLYQSK 10
DB      199 AEAETLYQTK 208
|||||

RESULT 5
US-10-094-749-2446
; Sequence 2446, Application US/10094749
; Patent No. 6979557
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2446
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2446

Query Match      87.2%; Score 41; DB 2; Length 608;
Best Local Similarity 80.0%; Pred. No. 6.2;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AEAESLYQSK 10
DB      412 AEAALYQTK 421
|||||

RESULT 6
US-09-949-016-10582
; Sequence 10582, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10582
```



;  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10582

Query Match 83.0%; Score 39; DB 2; Length 443;  
Best Local Similarity 80.0%; Pred. No. 11;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||:  
Db 236 AEASMYQIK 245

## RESULT 7

US-09-538-092-858  
; Sequence 858, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratSeqFormatter Version 0.9  
; SEQ ID NO 858  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P05787  
US-09-538-092-858

Query Match 83.0%; Score 39; DB 2; Length 482;  
Best Local Similarity 80.0%; Pred. No. 12;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||:  
Db 275 AEASMYQIK 284

## RESULT 8

US-09-919-497-79  
; Sequence 79, Application US/09919497  
; Patent No. 6773883  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 79  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-497-79

Query Match 83.0%; Score 39; DB 2; Length 483;  
Best Local Similarity 80.0%; Pred. No. 12;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||:  
Db 276 AEASMYQIK 285

## RESULT 9

US-09-949-016-11033  
; Sequence 11033, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11033  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-11033

Query Match 83.0%; Score 39; DB 2; Length 502;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEASLYQSK 10  
|||||:  
Db 295 AEASMYQIK 304

## RESULT 10

US-09-067-351-2  
; Sequence 2, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,351  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555

```
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 551 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
; CLONE: 2029060
; US-09-067-351-2

Query Match      80.9%; Score 38; DB 1; Length 551;
Best Local Similarity 80.0%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      1 AEAESLYQSK 10
Db      336 AEAESWYQTK 345

RESULT 11
US-09-360-490-2
; Sequence 2, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/360,490
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/067,351
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 551 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
; CLONE: 2029060
; US-09-360-490-2

Query Match      80.9%; Score 38; DB 2; Length 551;
Best Local Similarity 80.0%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      1 AEAESLYQSK 10
Db      336 AEAESWYQTK 345

; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 551 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
; CLONE: 2029060
; US-09-067-351-2

Query Match      80.9%; Score 38; DB 1; Length 551;
Best Local Similarity 80.0%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      1 AEAESLYQSK 10
Db      336 AEAESWYQTK 345

RESULT 12
US-09-949-016-6628
; Sequence 6628, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6628
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6628

Query Match      80.9%; Score 38; DB 2; Length 564;
Best Local Similarity 80.0%; Pred. No. 23;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      1 AEAESLYQSK 10
Db      350 AEAESWYQTK 359

RESULT 13
US-09-949-016-11035
; Sequence 11035, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11035
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-11035

Query Match      80.9%; Score 38; DB 2; Length 569;
Best Local Similarity 80.0%; Pred. No. 23;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      1 AEAESLYQSK 10
Db      355 AEAESWYQTK 364
```

## RESULT 14

US-09-949-016-11036  
; Sequence 11036, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C0001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 11036  
; LENGTH: 569  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-11036

Query Match 80.9%; Score 38; DB 2; Length 569;  
Best Local Similarity 80.0%; Pred. No. 23;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEAESLYQSK 10  
|||||:|  
Db 355 AEAESWYQTK 364

## RESULT 15

US-09-620-405B-487  
; Sequence 487, Application US/09620405B  
; Patent No. 6528054  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Yugui  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Hepler, William T.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; DIAGNOSIS OF BREAST CANCER  
; FILE REFERENCE: 210121.470C8  
; CURRENT APPLICATION NUMBER: US/09/620,405B  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 495  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 487  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-620-405B-487

Query Match 78.7%; Score 37; DB 2; Length 235;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AEAESLYQSK 10  
|||||:|  
Db 21 AEAESWYRSK 30

Search completed: July 5, 2006, 19:24:26  
Job time : 22.5243 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 10.9223 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928A-8

Perfect score: 45

Sequence: 1 ATNAENEFV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.\*

1: Piri:\*

2: Piri2:\*

3: Piri3:\*

4: Piri4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	91.1	581	1 KRMS2	keratin, type II c
2	41	91.1	643	1 KRHU2	keratin 1, type II
3	36	80.0	324	2 A81894	probable fructose-
4	36	80.0	324	2 F81126	fructose-1,6-bisph
5	36	80.0	553	2 I59009	epidermal keratin
6	36	80.0	583	2 G81412	aspartate-tRNA lig
7	35	77.8	390	2 A60093	cytokeratin, type
8	35	77.8	480	2 A56694	keratin, type II,
9	35	77.8	513	2 S08381	keratin, 58K type
10	35	77.8	590	2 A29904	keratin 5, type II
11	34	75.6	341	2 B86792	hypothetical prote
12	34	75.6	659	1 A64228	DNA ligase (NAD) (
13	33	73.3	350	2 F71144	probable galactoki
14	33	73.3	384	2 I61769	keratin 6d, type I
15	33	73.3	461	2 A43782	keratin, type II -
16	33	73.3	469	2 S05602	keratin K7, type I
17	33	73.3	489	2 B24177	keratin, 55K type
18	33	73.3	502	2 D95855	probable ABC trans
19	33	73.3	564	1 KRHUBA	keratin 6a, type I
20	33	73.3	564	1 KRHUBB	keratin 6b, type I
21	33	73.3	564	2 I61770	keratin 6e, type I
22	33	73.3	564	2 I61768	keratin 6c, type I
23	33	73.3	564	2 I61771	keratin 6f, type I
24	33	73.3	629	2 A29666	keratin, 65K type
25	33	73.3	629	2 S42629	keratin K3 - rabbi
26	33	73.3	638	2 I53169	cytokeratin 2 - hu
27	33	73.3	1524	2 G84721	hypothetical prote
28	33	73.3	1750	2 G84649	hypothetical prote
29	32	71.1	217	2 S45023	Grp-binding protei

Grp-binding protei  
hypothetical prote  
cytochrome-c oxida  
cytochrome-c oxida  
thioredoxine reduc  
probable type I re  
keratin, 64K type  
keratin, 64K type  
gamma-aminobutyrat  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
Grp-binding protei  
Grp-binding protei

ALIGNMENTS

RESULT 1

KRMS2

keratin, type II cytoskeletal - mouse (fragment)

N;Alternate names: 67-kDa type II keratin

C;Species: Mus musculus (house mouse)

C;Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004

C;Accession: A02951

R;Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.

J. Biol. Chem. 260, 7142-7149, 1985

A;Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000

ate filament subunits.

A;Reference number: A92535; MUID:85207740; PMID:2581964

A;Accession: A02951

A;Molecule type: mRNA

A;Residues: 1-581 <STE>

A;Cross-references: UNIPROT:P04104; UNIPARC:UPI0000173D5B; GB:M10937; NID:gl98622; PID:9

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil; intermediate filament

F;1-106/Domain: head (fragment) <HED>

F;1-70/Region: V1 subdomain (fragment)

F;71-106/Region: H1 subdomain

F;107-419/Domain: rod <ROD>

F;107-141/Region: coil 1A

F;142-153/Region: linker 1

F;154-254/Region: coil 1B

F;255-271/Region: linker 12

F;272-290/Region: coil 2A

F;291-298/Region: linker 2

F;299-419/Region: coil 2B

F;357/Region: stutter

F;420-581/Domain: tail <END>

F;420-439/Region: H2 subdomain

F;440-581/Region: V2 and E2 subdomains

Query Match 91.1%; Score 41; DB 1; Length 581;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9

Db 204 TNAENEFV 211

RESULT 2

KRHU2

keratin-1, type II, cytoskeletal - human

N;Alternate names: 67K type II epidermal keratin; cytokeratin 1

C;Species: Homo sapiens (man)

C;Date: 04-Dec-1986 #sequence\_revision 22-Oct-1999 #text\_change 10-Dec-1999

C;Accession: A22940; A02950; A43342

R;Johnson, L.D.; Idler, W.W.; Zhou, X.M.; Roop, D.R.; Steinert, P.M.

Proc. Natl. Acad. Sci. U.S.A. 82, 1896-1900, 1985

A:Reference number: A22940; MUID:85166239; PMID:2580302  
A:Accession: A22940  
A:Molecule type: DNA  
A:Residues: 1-643 <OH>  
A:Cross-references: UNIPARC:UPI0000173D59; GB:M98776; GB:M11215; GB:M11845; GB:M11846; N  
R:Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.  
J. Biol. Chem. 260, 7142-7149, 1985  
A:Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000  
late filament subunits  
A:Reference number: A92535; MUID:85207740; PMID:2581964  
A:Accession: A02950  
A:Molecule type: mRNA  
A:Residues: 151-183, 'K', 185-199, 'M', 201-204, 'K', 206-236, 'S', 238-239, 'R', 241-356, 'Y', 358-  
'S', 638-643 <STE>  
A:Cross-references: UNIPARC:UPI000016ABD0; GB:M10938; NID:gl86787; PIDN:AAA36153.1; PID:  
A:Experimental source: tissue neonatal foreskin  
R:Chipev, C.C.; Korge, B.P.; Markova, N.; Bale, S.J.; Digiovanna, J.J.; Compton, J.G.; S  
Cell 70, 821-828, 1992  
A:Title: A leucine---proline mutation in the H1 subdomain of keratin 1 causes epidermol  
A:Reference number: A43342; MUID:92386601; PMID:1381288  
A:Accession: A43342  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 144-146, 'P', 148-159, 'P', 161-183, 'K', 185-186 <CHI>  
A:Cross-references: UNIPARC:UPI0000173D5A; GB:M98776; GB:M11215; GB:M11845; GB:M11846; N  
A:Note: sequence extracted from NCBI backbone (NCBI:112784)  
C:Comment: The cytoskeletal and microfibrillar keratins are classified into two types, b  
atin I protein subunit appears to be a heterotetramer of two type I and two type II pro  
C:Comment: Keratin 1 is expressed in terminally differentiating epidermis.  
C:Genetics:  
A:Gene: GDB:KRT1  
A:Cross-references: GDB:128198; OMIM:139350  
A:Map position: 12ql1-12ql3  
A:Note: defects in this gene may result in epidermolytic hyperkeratosis  
C:Complex: heterotetramer of two type I, usually keratin 10 (see FIR:KRHU0), and two typ  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; heterotetramer; intermediate filament  
F:4-179/Domain: head <HD>  
F:4-143/Region: E1 and V1 subdomains  
F:14-179/Region: H1 subdomain  
F:180-492/Domain: rod <ROD>  
F:180-214/Region: coil 1A  
F:215-226/Region: linker 1  
F:227-327/Region: coil 1B  
F:328-344/Region: linker 12  
F:345-363/Region: coil 2A  
F:364-371/Region: linker 2  
F:372-492/Region: coil 2B  
F:430/Region: stutter  
F:493-643/Domain: tail <END>  
F:493-512/Region: H2 subdomain  
F:513-643/Region: V2 and E2 subdomains

Query Match 91.1%; Score 41; DB 1; Length 643;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9  
|||||||  
Db 277 TNAENEFV 284

RESULT 3  
A:Accession: A81894  
A:Title: probable fructose-bisphosphatase (EC 3.1.3.11) NMA1259 [imported] - Neisseria meningitidis  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C:Accession: A81894  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000

Query Match 80.0%; Score 36; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9  
|||||||  
Db 175 NAENEFV 181

RESULT 4  
F81126  
fructose-1,6-bisphosphatase NMB1060 [imported] - Neisseria meningitidis (strain MC58 sero  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: F81126  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: F81126  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <TET>  
A:Cross-references: UNIPROT:Q9JZH1; UNIPARC:UPI00000C4630; GB:AE002456; GB:AE002098; NID:  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1060  
C:Superfamily: fructose-1,6-bisphosphatase

Query Match 80.0%; Score 36; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9  
|||||||  
Db 175 NAENEFV 181

RESULT 5  
I59009  
epidermal keratin subunit II - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: I59009; A61205  
R:Steinert, P.M.; Parry, D.A.D.; Racoosin, E.L.; Idler, W.W.; Steven, A.C.; Trus, B.L.; F  
Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713, 1984  
A:Title: The complete cDNA and deduced amino acid sequence of a type II mouse epidermal  
A:Reference number: I59009; MUID:85014838; PMID:6207530  
A:Accession: I59009  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-553 <RES>  
A:Cross-references: UNIPROT:P50446; UNIPARC:UPI000016CE87; GB:K02108; NID:g198634; PIDN:  
R:Finch, J.; Andrews, K.; Krieg, P.; Fuerstenberger, G.; Slaga, T.; Ootsuyama, A.; Tanoo  
Carcinogenesis 12, 1519-1522, 1991  
A:Title: Identification of a cloned sequence activated during multi-stage carcinogenesis  
A:Reference number: A61205; MUID:91316763; PMID:1713533

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: A81894  
A:Molecule type: DNA  
A:Residues: 1-324 <PAR>  
A:Cross-references: UNIPROT:Q9JUL6; UNIPARC:UPI00000C4B5F; GB:AL162755; GB:AL157959; NID:  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: fbp; NMA1259  
C:Superfamily: fructose-1,6-bisphosphatase  
C:Keywords: phosphoric monoester hydrolase

Query Match 80.0%; Score 36; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9  
|||||||  
Db 175 NAENEFV 181

RESULT 4  
F81126  
fructose-1,6-bisphosphatase NMB1060 [imported] - Neisseria meningitidis (strain MC58 sero  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: F81126  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: F81126  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <TET>  
A:Cross-references: UNIPROT:Q9JZH1; UNIPARC:UPI00000C4630; GB:AE002456; GB:AE002098; NID:  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1060  
C:Superfamily: fructose-1,6-bisphosphatase

Query Match 80.0%; Score 36; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9  
|||||||  
Db 175 NAENEFV 181

RESULT 5  
I59009  
epidermal keratin subunit II - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: I59009; A61205  
R:Steinert, P.M.; Parry, D.A.D.; Racoosin, E.L.; Idler, W.W.; Steven, A.C.; Trus, B.L.; F  
Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713, 1984  
A:Title: The complete cDNA and deduced amino acid sequence of a type II mouse epidermal  
A:Reference number: I59009; MUID:85014838; PMID:6207530  
A:Accession: I59009  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-553 <RES>  
A:Cross-references: UNIPROT:P50446; UNIPARC:UPI000016CE87; GB:K02108; NID:g198634; PIDN:  
R:Finch, J.; Andrews, K.; Krieg, P.; Fuerstenberger, G.; Slaga, T.; Ootsuyama, A.; Tanoo  
Carcinogenesis 12, 1519-1522, 1991  
A:Title: Identification of a cloned sequence activated during multi-stage carcinogenesis  
A:Reference number: A61205; MUID:91316763; PMID:1713533

A;Accession: A61205  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 528-553 <FIN>  
A;Cross-references: UNIPARC:UPI00001774C8  
C;Genetics:  
A;Gene: KER2  
C;Superfamily: cytoskeletal keratin

Query Match 80.0%; Score 36; DB 2; Length 553;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TNAENEFV 9  
|:|||||  
Db 250 TDAENEFV 257

RESULT 6  
G81412  
aspartate-tRNA ligase (EC 6.1.1.12) Cj0640c [imported] - Campylobacter jejuni (strain NC  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: G81412  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: G81412  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-583 <PAR>  
A;Cross-references: UNIPROT:Q9PHM7; UNIPARC:UPI000013638A; GB:AL1139075; GB:AL111168; NID  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: aspS; Cj0640C  
C;Superfamily: lysine-tRNA ligase  
C;Keywords: ligase

Query Match 80.0%; Score 36; DB 2; Length 583;  
Best Local Similarity 77.8%; Pred. No. 14;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATNAENEFV 9  
|:|||||  
Db 61 ASNARNEFV 69

RESULT 7  
A60093  
cytokeratin, type II, early ectodermal - chicken (fragment)  
C;Species: Gallus gallus (chicken)  
C;Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 30-Sep-1993  
C;Accession: A60093  
R;Charlebois, T.S.; Spencer, D.H.; Tarkington, S.K.; Henry, J.J.; Grainger, R.M.  
Development 108, 33-45, 1990  
A;Title: Isolation of a chick cytokeratin cDNA clone indicative of regional specializat  
A;Reference number: A60093; MUID:90276238; PMID:1693557  
A;Accession: A60093  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-390 <CHA>  
A;Cross-references: UNIPARC:UPI00001774C6  
C;Superfamily: cytoskeletal keratin  
C;Keywords: intermediate filament

Query Match 77.8%; Score 35; DB 2; Length 390;  
Best Local Similarity 87.5%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TNAENEFV 9  
|:|||||

Db 108 TTAENEFV 115

RESULT 8  
A56694  
keratin, type II, KRT2.13 - sheep  
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C;Date: 21-Jul-1995 #sequence\_revision 28-Jul-1995 #text\_change 09-Jul-2004  
C;Accession: A56694  
R;Powell, B.C.; Crocker, L.A.; Rogers, G.E.  
DNA Seq. 3, 401-405, 1993  
A;Title: Complete sequence of a hair-like intermediate filament type II keratin gene.  
A;Reference number: A56694; MUID:94033628; PMID:7693053  
A;Accession: A56694  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-480 <POW>  
A;Cross-references: UNIPROT:Q06018; UNIPARC:UPI0000088521; GB:X72379; NID:g297105; PIDN:  
C;Genetics:  
A;Gene: KRT2.13  
A;Introns: 140/3; 210/2; 230/3; 262/3; 317/3; 359/3; 433/2; 444/1  
C;Superfamily: cytoskeletal keratin

Query Match 77.8%; Score 35; DB 2; Length 480;  
Best Local Similarity 87.5%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TNAENEFV 9  
|:|||||  
Db 219 TTAENEFV 226

RESULT 9  
S08381  
keratin, 58K type II, cytoskeletal (clone pXenCK55(5/6)) - African clawed frog  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 09-Jul-2004  
C;Accession: S08381  
R;Fouquet, B.; Herrmann, H.; Franz, J.K.; Franke, W.W.  
Development 104, 533-548, 1988  
A;Title: Expression of intermediate filament proteins during development of Xenopus lae  
A;Reference number: S08381; MUID:90032362; PMID:2478354  
A;Accession: S08381  
A;Molecule type: mRNA  
A;Residues: 1-513 <FOU>  
A;Cross-references: UNIPROT:P16878; UNIPARC:UPI000017151D; EMBL:X14427; NID:g64625; PIDN:  
C;Superfamily: cytoskeletal keratin

Query Match 77.8%; Score 35; DB 2; Length 513;  
Best Local Similarity 87.5%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TNAENEFV 9  
|:|||||  
Db 225 TGAENEFV 232

RESULT 10  
A29904  
keratin 5, type II, epidermal - human  
N;Alternate names: 58K type II keratin; cytokeratin 5; K5  
C;Species: Homo sapiens (man)  
C;Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 09-Jul-2004  
C;Accession: A29904; A60865; A32568; A27679; I58080  
R;Eckert, R.L.; Rorke, E.A.  
DNA 7, 337-345, 1988  
A;Title: The sequence of the human epidermal 58-kD (5) type II keratin reveals an absenc  
A;Reference number: A29904; MUID:88296082; PMID:2456903  
A;Accession: A29904  
A;Molecule type: DNA; mRNA  
A;Residues: 1-590 <ECK>  
A;Cross-references: UNIPROT:P13647; UNIPARC:UPI000012DB0D; GB:M21389; NID:g186697; PIDN:  
R;Galup, C.; Darmon, M.Y.

J. Invest. Dermatol. 91, 39-42, 1988  
A:Title: Isolation and characterization of a cDNA clone coding for human epidermal keratin A;Reference number: A60865; MUID:88258089; PMID:2455002  
A:Accession: A60865  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 348-374, 'E', 376-386, 'S', 388-436, 438-542, 'S', 544-590 <GAL>  
A:Cross-references: UNIPARC:UPI00001774BD  
R:Lersch, R.; Stellmach, V.; Stocks, C.; Gludice, G.; Fuchs, E.  
Mol. Cell. Biol. 9, 3685-3697, 1989  
A:Title: Isolation, sequence, and expression of a human keratin K5 gene: transcriptional control of keratin K5 gene; A:Reference number: A32568; MUID:89384595; PMID:2476664  
A:Accession: A32568  
A:Molecule type: DNA  
A:Residues: 2-8, 'SGA', 12-36, 'GP', 39-78, 'S', 80-196, 'E', 198-260, 'Q', 262-270, 'H', 272-386, 'S', 388-436, 'E', 438-542, 'S', 544-590 <GAL>  
A:Cross-references: UNIPARC:UPI00001774BE; GB:M28496; NID:g186728  
A>Note: the authors translated the codon CGR for residue 15 as Pro and GAG for residue 2  
R:Lersch, R.; Fuchs, E.  
Mol. Cell. Biol. 8, 486-493, 1988  
A:Title: Sequence and expression of a type II keratin, K5, in human epidermal cells.  
A:Reference number: A27679; MUID:88094424; PMID:2447486  
A:Accession: A27679  
A:Molecule type: mRNA  
A:Residues: 83-196, 'E', 198-260, 'Q', 262-270, 'H', 272-386, 'S', 388-557, 'S', 559-590 <LE2>  
A:Cross-references: UNIPARC:UPI000016ABAG; GB:M19723; NID:g186726; PIDN:AAA36145.1; PID: R:Lane, E.B.; Rugg, E.L.; Navsaria, H.; Leigh, I.M.; Hegarty, A.H.M.; Ishida-Yamamoto, Nature 356, 244-246, 1992  
A:Title: A mutation in the conserved helix termination peptide of keratin 5 in hereditary ichthyosis; A:Reference number: I58080; MUID:92204232; PMID:1372711  
A:Accession: I58080  
A:Molecule type: mRNA  
A>Status: translated from GB/EMBL/DDBT  
A:Residues: 467-474, 'G', 476-481 <LAN>  
A:Cross-references: UNIPARC:UPI000011F7CB; GB:S90014; NID:g247707; PIDN:AAB21859.1; PID: C:Comment: This type II keratin is expressed together with the type I keratin 14 in the C:Genetics:  
A:Gene: GDB:KRT5; EBS2  
A:Cross-references: GDB:128110; OMIM:148040  
A:Map position: 12q11-12q13  
A:Introns: 185/3; 257/2; 277/3; 309/3; 364/3; 406/3; 480/2; 492/1  
A>Note: defects in this gene may result in Dowling-Meara type epidermolysis bullosa simplex; C:Complex: heterotetramer of two type I, usually keratin 14 (see PIR:KRHUE), and two type C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; cytoskeleton; epidermis; heterotetramer

Query Match 77.8%; Score 35; DB 2; Length 590;  
Best Local Similarity 87.5%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TNAENEV 9  
266 TTAENEV 273

Db

RESULT 11  
B86792  
hypothetical protein gpda [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 05-Oct-2004  
C:Accession: B86792  
R:Belotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. Genome Res. 11, 731-753, 2001  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: B86792  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-341 <STO>  
A:Cross-references: UNIPROT:Q9CFX6; UNIPARC:UPI000012B98F; GB:AE005176; PID:g12724320; F  
C:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: gpda

C:Superfamily: glycerol-3-phosphate dehydrogenase (NAD)

Query Match 75.6%; Score 34; DB 2; Length 341;  
Best Local Similarity 75.0%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TNAENEV 9  
331 TTAENEV 338

Db

RESULT 12  
A64228  
DNA ligase (NAD) (EC 6.5.1.2) - Mycoplasma genitalium  
C:Species: Mycoplasma genitalium  
C>Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 31-Dec-2004  
C:Accession: A64228  
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.N.; C.A.; Venter, J.C.  
Science 270, 397-403, 1995  
A:Title: The minimal gene complement of Mycoplasma genitalium.  
A:Reference number: A64200; MUID:96026346; PMID:7569993  
A:Accession: A64228  
A:Molecule type: DNA  
A:Residues: 1-659 <TIGR>  
A:Cross-references: UNIPROT:P47496; UNIPARC:UPI0000129683; GB:U39703; GB:I43967; NID:g104  
A:Experimental source: strain G-37  
C:Genetics:  
A:Genetic code: SGC3  
A:Start codon: GTG  
C:Superfamily: DNA ligase (NAD), Liga type  
C:Keywords: ligase; NAD

Query Match 75.6%; Score 34; DB 1; Length 659;  
Best Local Similarity 75.0%; Pred. No. 42;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEV 9  
102 TNSKNEV 109

Db

RESULT 13  
F71144  
probable galactokinase - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Oct-2004  
C:Accession: F71144  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon; A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: F71144  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-350 <KAW>  
A:Cross-references: UNIPROT:O58107; UNIPARC:UPI000012B05F; GB:AP000002; NID:g3236129; PFI  
A:Experimental source: strain OT3  
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0369  
C:Superfamily: galactokinase

Query Match 73.3%; Score 33; DB 2; Length 350;  
Best Local Similarity 77.8%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 9  
133 AKKAENEV 141

Db



## RESULT 14

I61769  
keratin 6d, type II - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
C:Accession: I61769  
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A:Reference number: A57398; MUID:95355491; PMID:7543104  
A:Accession: I61769  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-384 <RES>  
A:Cross-references: UNIPROT:P48667; UNIPARC:UPI000012DB17; GB:L42602; NID:g914823; GB:L4  
28; GB:L42608; NID:g914829; GB:L42609; NID:g914830; GB:L42610; NID:g1488252; PIDN:AAB606  
C:Genetics:  
A:Gene: KRT6D  
A:Introns: 72/2; 92/3; 124/3; 179/3; 221/3; 295/2; 307/1  
A:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 73.3%; Score 33; DB 2; Length 384;  
Best Local Similarity 87.5%; Pred.No. 38;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TNAENEVF 9

Db 81 TAAENEVF 88

## RESULT 15

A43782  
keratin, type II - eastern newt (fragment)  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C>Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A43782  
R:Ferretti, P.; Brookes, J.P.; Brown, R.  
Development 111, 497-507, 1991  
A:Title: A new type II keratin restricted to normal and regenerating limbs and tails is  
A:Reference number: A43782; MUID:91372147; PMID:1716554  
A:Accession: A43782  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-461 <PER>  
A:Cross-references: UNIPROT:O42435; UNIPARC:UPI00001774C4; GB:X57671  
C:Superfamily: cytoskeletal keratin

Query Match 73.3%; Score 33; DB 2; Length 461;  
Best Local Similarity 87.5%; Pred.No. 46;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TNAENEVF 9

Db 142 TAAENEVF 149

Search completed: July 5, 2006, 19:20:29  
Job time : 11.9223 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 92.0971 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-8  
Perfect score: 45  
Sequence: 1 ATNAENEFV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 7.2.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	91.1	619	2	Q6E1Y9 CANFA
2	41	91.1	625	2	Q6IMF3 RAT
3	41	91.1	627	1	K2C1 MOUSE
4	41	91.1	637	2	Q9D2K8 MOUSE
5	41	91.1	643	1	K2C1 HUMAN
6	38	84.4	222	2	Q4L0E7 MOUSE
7	38	84.4	224	2	Q678L1 MOUSE
8	38	84.4	572	1	K2C1B MOUSE
9	38	84.4	869	2	Q63ER5_BACCCZ
10	38	84.4	869	2	Q81U13_BACAN
11	36	80.0	207	2	Q5H887_NEIGO
12	36	80.0	324	2	Q9JUL6_NEIMA
13	36	80.0	324	2	Q9ZH1_NEIMEB
14	36	80.0	324	2	Q5F8C3_NEIGI
15	36	80.0	529	2	Q919P5_RANCA
16	36	80.0	557	2	Q6IG07 RAT
17	36	80.0	580	2	Q98UJ3_XENLA
18	36	80.0	583	1	SYD_CAMJF
19	36	80.0	583	1	SYD_CAMJR
20	36	80.0	590	2	Q5XGT4_XENLA
21	36	80.0	598	2	Q701L6_HUMAN
22	36	80.0	600	1	KRHB4_HUMAN
23	36	80.0	600	2	Q6ISB0_HUMAN
24	36	80.0	603	1	KRHB4_MOUSE
25	35	77.8	241	2	Q4Y317_PLACH
26	35	77.8	375	2	Q2UTZ4_ASPOR
27	35	77.8	381	2	Q32MB2_HUMAN
28	35	77.8	480	2	Q06018_SHEEP
29	35	77.8	492	1	K2CO_CHICK
30	35	77.8	512	1	K2C5_XENLA
31	35	77.8	519	1	K2C1E_RAT

32	35	77.8	529	2	Q86V45_HUMAN
33	35	77.8	539	2	Q6NXH9_MOUSE
34	35	77.8	540	2	Q86Y46_HUMAN
35	35	77.8	543	2	Q7RTS7_HUMAN
36	35	77.8	553	2	Q6IG03_RAT
37	35	77.8	554	2	Q3UV11_MOUSE
38	35	77.8	561	1	K2C6B_MOUSE
39	35	77.8	576	1	K2C5_RAT
40	35	77.8	580	1	K2C5_MOUSE
41	35	77.8	581	2	Q6IFZ5_RAT
42	35	77.8	590	1	K2C5_HUMAN
43	35	77.8	599	2	Q6PVZ5_CHICK
44	35	77.8	601	1	K2C5_BOVIN
45	35	77.8	606	2	Q32P04_MOUSE

Q86V45	homo sapien
Q6NXH9	mus musculus
Q86Y46	homo sapien
Q7RTS7	homo sapien
Q6IG03	rattus norv
Q3UV11	mus musculus
Q92331	mus musculus
Q6P642	rattus norv
Q92242	mus musculus
Q6IFZ5	rattus norv
P13647	homo sapien
Q6PVZ5	gallus gall
Q5XQ05	bos taurus
Q32P04	mus musculus

#### ALIGNMENTS

##### RESULT 1

Q6E1Y9\_CANFA  
ID Q6E1Y9\_CANFA PRELIMINARY; PRT; 619 AA.  
AC Q6E1Y9;  
DT 16-AUG-2004, integrated into UniProtKB/TREMBL.  
DT 16-AUG-2004, sequence version 1.  
DT 07-FEB-2006, entry version 10.  
DE Epithelial keratin 1.  
GN Name=Ker1;  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
OC Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15627753; DOI=10.1159/000081527;  
RA Credille K.M., Guyon R., Andre C., Murphy K., Tucker K.,  
RA Barnhart K.F., Dunstan R.W.;  
RT "Comparative sequence analysis and radiation hybrid mapping of two  
RT epidermal type II keratin genes in the dog: keratin I and keratin  
RT 2e.";  
RL Cytogenet. Genome Res. 108:328-332(2005).  
CC -|- SIMILARITY: Belongs to the intermediate filament family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC -----  
DR EMBL; AV318945; AAQ83910.1; -; mRNA.  
DR Ensembl; ENSCAFG00000007212; Canis familiaris.  
DR GO; GO:0005982; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin I.  
DR InterPro; IPR003054; Keratin II.  
DR PANTHER; PTHR18993.SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PRO1248; TYPE1KERATIN.  
DR PRINTS; PRO1276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 619 AA; 63790 MW; B45AB8B709AD0909F CRC64;

Query Match 91.1%; Score 41; DB 2; Length 619;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TNAENEFV 9

Db 279 TNAENEFV 286

##### RESULT 2

Q6IMF3\_RAT

```

ID O6IME3_RAT PRELIMINARY; PRT; 625 AA.
AC O6IME3;
DT 05-JUL-2004, integrated into UniProtKB/TREMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Type II keratin Kbl.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;
RA Hesse M., Zimek A., Weber K., Magin T.M.;
RT "Comprehensive analysis of keratin gene clusters in humans and
rodents."
RL Eur. J. Cell Biol. 83:19-26(2004).
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
EMBL/GenBank/DBJ third party annotation (TPA) entry.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BK001580; DAA02055.1; -; mRNA.
DR Ensembl; ENSRNOG0000028996; Rattus norvegicus.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin I.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR18893.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 625 AA; 64831 MW; 790AAC59E2A2707B CRC64;

Query Match 91.1%; Score 41; DB 2; Length 625;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9
Db 277 TNAENEFV 284
|||||||

RESULT 3
K2C1_MOUSE STANDARD; PRT; 627 AA.
AC P04104;
DT 01-NOV-1986, integrated into UniProtKB/Swiss-Prot.
DT 30-MAY-2000, sequence version 2.
DT 07-FEB-2006, entry version 50.
DE Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1)
DE (K1) (67 kDa cyokeratin)
GN Name=Krt1; Synonyms=Krt2-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=85207740; PubMed=2581964;
RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
RA Roop D.R.;
RT "Amino acid sequences of mouse and human epidermal type II keratins of
Mr 67,000 provide a systematic basis for the structural and functional
diversity of the end domains of keratin intermediate filament
subunits";

```

```

RL J. Biol. Chem. 260:7142-7149(1985).
RN [2]
RP SEQUENCE REVISION.
RA Roop D.R.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: Heterotetramer of two type I and two type II keratins.
CC -!- PTM: Undergoes deamination of some arginine residues
(citrullination) (By similarity).
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and
microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
55 and 56-70 kilodaltons, respectively).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; M10937; AAD05191.1; -; mRNA.
DR PIR; A02951; KRMS2.
DR HSSP; P08670; 1GK7.
DR SWISS-2DPAGE; P04104; MOUSE.
DR Ensembl; ENSMUSG00000046834; Mus musculus.
DR MGI; MGI:96698; Krt2-1.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin I.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR18893.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Citrullination; Coiled coil; Intermediate filament; Keratin.
FT INIT MET 0 0 By similarity.
FT CHAIN 1 627 Keratin type II cytoskeletal 1.
FT REGION 1 186 Head.
FT REGION 187 496 Rod.
FT REGION 187 222 Coil 1A.
FT REGION 223 242 Linker 1.
FT REGION 243 333 Coil 1B.
FT REGION 334 357 Linker 12.
FT REGION 358 496 Coil 2.
FT REGION 497 627 Tail.
FT SITE 451 451 Stutter.
SQ SEQUENCE 627 AA; 65092 MW; EF7E848654539578 CRC64;

Query Match 91.1%; Score 41; DB 1; Length 627;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9
Db 285 TNAENEFV 292
|||||||

RESULT 4
Q9D2K8_MOUSE PRELIMINARY; PRT; 637 AA.
ID Q9D2K8_MOUSE
AC Q9D2K8;
DT 01-JUN-2001, integrated into UniProtKB/TREMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 23.
DE 0 day neonate head cDNA, RIKEN full-length enriched library,
DE clone:4833436C19 product:keratin complex 2, basic, gene 1, full insert
DE sequence.
GN Name=Krt2-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=C57BL/6J; TISSUE=Head;

```

RA MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44 (1999).  
[2]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Head;  
RC PubMed=16141072; DOI=10.1126/science.1112014;  
RX Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impiombato A., Apweiler R.N., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Cluttenbuck D.R.,  
RA Crowe M.L., Daila E., Dallymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Fukaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Farar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sander A., Schneider C., Sheng Y.,  
RA Schonbach C., Sekiguchi K., Sempile C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoji K., Tan S.L., Tang S., Taylor J.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Walestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tisham M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563 (2005).  
[3]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Head;  
RC PubMed=16141073; DOI=10.1126/science.1112009;  
RX RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the PANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566 (2005).  
[4]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami H., Schonbach C., Gojobori T.,  
RA Schirral L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perle G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,

RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempile C.A., Secou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Walestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RT Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
[5]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RX Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690 (2001).  
[6]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;  
RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630 (2000).  
[7]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RX Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitesunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771 (2000).  
[8]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Head;  
RC Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,

RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
RA Tejina Y., Tova T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
RA Muramatsu M., Hayashizaki Y.,  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -/- SIMILARITY: Belongs to the intermediate filament family.  
CC -----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: AK019521; BAB31776.1; -, mRNA.  
DR HSSP: P08670; 1GK7.  
DR Ensemble: ENSMUSG0000046834; Mus musculus.  
DR MGI: MGI:96698; Krt2-1.  
DR GO: GO:0005982; C:intermediate filament; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR InterPro: IPR001664; IF.  
DR InterPro: IPR002957; Keratin I.  
DR InterPro: IPR003054; Keratin II.  
DR PANTHER: PTHR18993.SF5; Keratin\_II; 1.  
DR Pfam: PF00038; Filament; 1.  
DR PRINTS: PR01248; TYPE1KERATIN.  
DR PRINTS: PR01276; TYPE2KERATIN.  
DR PROSITE: PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 637 AA; 65627 MW; D6344FDD0468ECA7 CRC64;  
  
Query Match 91.1%; Score 41; DB 2; Length 637;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 TNAENEFV 9  
DB 286 TNAENEFV 293  
|||||||  
  
RESULT 5  
K2C1 HUMAN STANDARD; PRT; 643 AA.  
AC P04264; Q14720; Q6GSJ0; Q9H298;  
DT 20-MAR-1987, integrated into UniprotKB/Swiss-Prot.  
DT 01-FEB-1996, sequence version 4.  
DT 07-FEB-2006, entry version 76.  
DE Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1)  
DE (K1) (67 kDa cytokeratin) (Hair alpha protein).  
GN Name=KRT1; Synonyms=KRTA;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RX MEDLINE=85166239; PubMed=2580302;  
RA Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;  
RA "Structure of a gene for the human epidermal 67-kDa keratin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RA Whitlock N.V., Eady R.A.J., McGrath J.A.;  
RA "Genomic organization of the human keratin 1 gene.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RA Hatzell S.J., Eady R.A.J., Wennerstrand L., Dopping-Hepenstal P.J.,  
RA Leigh I.M., Munro C., Kelsell D.P.;  
RA "Novel splice site mutation in keratin 1 underlies mild epidermolytic  
RT palmoplantar keratoderma in three kindreds.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA], AND VARIANT TYR-357.  
RC TISSUE=Skin;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards A., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP PRELIMINARY NUCLEOTIDE SEQUENCE [MRNA] OF 151-643.  
RX MEDLINE=85207740; PubMed=2581964;  
RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,  
RA Roop D.R.;  
RT "Amino acid sequences of mouse and human epidermal type II keratins of  
RT Mr 67,000 provide a systematic basis for the structural and functional  
RT diversity of the end domains of keratin intermediate filament  
RL subunits.";  
RL J. Biol. Chem. 260:7142-7149(1985).  
RN [6]  
RP PROTEIN SEQUENCE OF 1-7.  
RC TISSUE=Platelet;  
RX MEDLINE=22608298; PubMed=12655801; DOI=10.1038/nbt810;  
RA Gevaert K., Goethals M., Martens L., Van Damme J., Staes A.,  
RA Thomas G.R., Vandekerckhove J.;  
RT "Exploring proteomes and analyzing protein processing by mass  
RT spectrometric identification of sorted N-terminal peptides.";  
RL Nat. Biotechnol. 21:566-569(2003).  
RN [7]  
RP CITRULLINATION  
RX MEDLINE=96374388; PubMed=8780679; DOI=10.1006/bbrc.1996.1240;  
RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;  
RT "Preferential deamination of keratin K1 and filaggrin during the  
RT terminal differentiation of human epidermis.";  
RL Biochem. Biophys. Res. Commun. 225:712-719(1996).  
RN [8]  
RP VARIANT EHK PRO-160.  
RX MEDLINE=92386601; PubMed=1381288; DOI=10.1016/0092-8674(92)90315-4;  
RA Chipav C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,  
RA Compton J.G., Steinert P.M.;  
RT "A leucine-->proline mutation in the H1 subdomain of keratin 1 causes  
RT epidermolytic hyperkeratosis.";  
RL Cell 70:821-828(1992).  
RN [9]  
RP VARIANT ALLELE 1B 559-GLY--TYR-565 DEL.  
RX MEDLINE=93107743; PubMed=1281859; DOI=10.1111/1523-1747.epi2614149;  
RA Korge B.P., Compton J.G., Steinert P.M., Mischke D.;  
RT "The two size alleles of human keratin 1 are due to a deletion in the  
RT glycine-rich carboxyl-terminal V2 subdomain.";  
RL J. Invest. Dermatol. 99:697-702(1992).  
RN [10]  
RP VARIANT EHK GLN-489.  
RX MEDLINE=92376531; PubMed=1380725;  
RA Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,  
RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;  
RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic  
RT hyperkeratosis.";  
RL Science 257:1128-1130(1992).  
RN [11]  
RP VARIANT EHK CYS-481.  
RX MEDLINE=94216497; PubMed=7512983;  
RA Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;  
RT "Genetic mutations in the K1 and K10 genes of patients with

epidermolytic hyperkeratosis. Correlation between location and disease severity." J. Clin. Invest. 93:1533-1542(1994).

[12]  
VARIANTS EHK GLY-154; SER-187 AND PRO-192.  
MEDLINE=94117869; PubMed=7507151; DOI=10.1111/1523-1747.ep12371725;  
Yang J.-M., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,  
Steinert P.M., Compton J.G., McGlothlin J.A., Griffiths W.A.D., Eady R.A.J.,  
"Mutations in the H1 and I2 domains in the keratin 1 gene in  
epidermolytic hyperkeratosis." Invest. Dermatol. 102:17-23(1994).

[13]  
VARIANTS EHK PRO-185 AND SER-187.  
MEDLINE=94117870; PubMed=7507152; DOI=10.1111/1523-1747.ep12371726;  
McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,  
Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,  
Morley S.M.,  
"Mutations in the rod 1A domain of keratins 1 and 10 in bullous  
congenital ichthyosiform erythroderma (BCIE)." J. Invest. Dermatol. 102:24-30(1994).

[14]  
VARIANT NEPPK ILE-73.  
MEDLINE=95096501; PubMed=7528239; DOI=10.1111/1523-1747.ep12412771;  
Kimonis V., Digiovanna J.J., Yang J.-M., Doyle S.Z., Bale S.J.,  
Compton J.G.,  
"A mutation in the V1 end domain of keratin 1 in non-epidermolytic  
palmar-plantar keratoderma." J. Invest. Dermatol. 103:764-769(1994).

[15]  
VARIANT EHK VAL-339.  
MEDLINE=99072666; PubMed=9856846;  
DOI=10.1046/j.1523-1747.1998.00389.x;  
Kremer H., Lavrijssen A.P., McLean W.H.I., Lane E.B., Melchers D.,  
Rutter D.J., Marimón E.C., Steijlen P.M.,  
"An atypical form of bullous congenital ichthyosiform erythroderma is  
caused by a mutation in the L12 linker region of keratin 1." J. Invest. Dermatol. 111:1224-1226(1998).

[16]  
VARIANTS AEI PHE-478 AND THR-478.  
MEDLINE=99162195; PubMed=10053007;  
Sybert V.P., Francis J.S., Corden L.D., Smith L.T., Weaver M.,  
Stephens K., McLean W.H.I.,  
"Cyclic ichthyosis with epidermolytic hyperkeratosis: a phenotype  
conferred by mutations in the 2B domain of keratin K1." Am. J. Hum. Genet. 64:732-738(1999).

[17]  
VARIANT EHK THR-187.  
MEDLINE=99247363; PubMed=10232403;  
Arin M.J., Longley M.A., Kuster W., Huber M., Hohl D., Rothnagel J.A.,  
Roop D.R.,  
"An asparagine to threonine substitution in the 1A domain of keratin  
1: a novel mutation that causes epidermolytic hyperkeratosis." Exp. Dermatol. 8:124-127(1999).

[18]  
VARIANT AEI PHE-478.  
MEDLINE=20062276; PubMed=10597140;  
Michael E.J., Schneiderman P., Grossman M.E., Christiano A.M.,  
"Epidermolytic hyperkeratosis with polycyclic psoriasiform plaques  
resulting from a mutation in the keratin 1 gene." Exp. Dermatol. 8:501-503(1999).

[19]  
VARIANT EHK PRO-213.  
MEDLINE=20305470; PubMed=10844506;  
DOI=10.1046/j.1365-2230.2000.00625.x;  
Cserhalmi-Friedman P.B., Squeo R., Gordon D., Garzon M.,  
Schneiderman P., Grossman M.E., Christiano A.M.,  
"Epidermolytic hyperkeratosis in a Hispanic family resulting from a  
mutation in the keratin 1 gene." Clin. Exp. Dermatol. 25:241-243(2000).

[20]  
VARIANT EHK THR-478.  
MEDLINE=20151078; PubMed=10688370;  
DOI=10.1034/j.1600-0625.2000.009001016.x;

RA Arin M.J., Longley M.A., Epstein E.H. Jr., Rothnagel J.A., Roop D.R.;  
RT "Identification of a novel mutation in keratin 1 in a family with  
RL epidermolytic hyperkeratosis." Exp. Dermatol. 9:16-19(2000).

RN [21]  
RP VARIANT EHK ASP-154.  
RX MEDLINE=21423058; PubMed=11531804;  
DOI=10.1046/j.1365-2133.2001.04327.x;  
RA Whitlock N.V., Ashton G.H.S., Griffiths W.A.D., Eady R.A.J.,  
McGrath J.A.,  
RT "New mutations in keratin 1 that cause bullous congenital  
RT ichthyosiform erythroderma and keratin 2e that cause ichthyosis  
RT bullosa of Siemens." Br. J. Dermatol. 145:330-335(2001).

RN [22]  
RP INVOLVEMENT IN IHCM.  
RX MEDLINE=21184521; PubMed=11286616;  
DOI=10.1046/j.1523-1747.2001.01292.x;

Query Match 91.1%; Score 41; DB 1; Length 643;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TNAENFV 9  
Db 277 TNAENFV 284  
|||||||

RESULT 6  
Q4L0E7 MOUSE PRELIMINARY; PRT; 222 AA.  
ID Q4L0E7 MOUSE  
AC Q4L0E7  
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 02-AUG-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Type II cytokeratin Kb39 (Fragment).  
GN Name=4732484G22Rik;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=NMRI; TISSUE=Tail skin;  
RA Zimek A., Hesse M.;  
RT "Cornea-specific cytochrome Kb3 seems to be absent in the rodent  
RT lineage." Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: AY651826; AAV85505.1; -; mRNA.  
DR MGI: 3588209; 4732484G22Rik.  
DR GO: GO:0005882; C:intermediate filament; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR InterPro: IPR002957; Keratin I.  
DR InterPro: IPR003054; Keratin II.  
DR PANTHER: PTHR1893:SP5; Keratin II; 1.  
DR Pfam: PF00038; Filament; 1.  
DR PRINTS: PR01248; TYPE1KERATIN.  
DR PRINTS: PR01276; TYPE2KERATIN.  
KW Keratin.  
FT NON\_TER 1 222  
FT NON\_TER 222 222  
SQ SEQUENCE 222 AA; 25801 MW; 483B4F589032C1D1 CRC64;  
Query Match 84.4%; Score 38; DB 2; Length 222;  
Best Local Similarity 87.5%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TNAENFV 9

```

Db          78 TNAENDFV 85
|||||:|
RESULT 7
Q678L1_MOUSE
ID Q678L1_MOUSE PRELIMINARY; PRT; 224 AA.
AC Q678L1_MOUSE
DT 11-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 12.
DE Keratin 1b (Fragment).
GN Name=4732484G22Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MRI; TISSUE=Tail skin;
RA Zimek A., Hesse M., Magin T., Weber K.;
RT "Homologies between the human, mouse and rat keratin-clusters.";
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY379112; AAR23390.1; -; mRNA.
DR Ensembl; ENSMUSG0000067594; Mus musculus.
DR MGI; MGI:3588209; 4732484G22Rik.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR18993:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
FT NON_TER 1 1
FT NON_TER 224 224
SQ SEQUENCE 224 AA; 25995 MW; DD6F783B4F589032 CRC64;

Query Match 84.4%; Score 38; DB 2; Length 224;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEV 9
|||||:|
Db 78 TNAENDFV 85

RESULT 8
K2C1B_MOUSE
ID K2C1B_MOUSE STANDARD; PRT; 572 AA.
AC Q61F26;
DT 24-MAY-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DE Keratin, type II cytoskeletal 1b.
DE Keratin, type II cytoskeletal 1b (Type II keratin K539) (Embryonic
DE type II keratin-1).
GN Name=Krt1b;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG The mouse genome sequencing consortium;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

```

```

RN [2]
RP PARTIAL PROTEIN SEQUENCE, IDENTIFICATION, AND CITRULLINATION.
RC STRAIN=C57BL/6; TISSUE=Epidermis;
RX PubMed=15619433; DOI=10.1016/j.jdermsci.2004.07.003;
RA Senhu T., Ishida-Yamamoto A., Takahashi H., Iizuka H.;
RT "Prediction of a coding sequence for a novel type II keratin from N-
RT terminal sequences of mouse epidermal proteins site-specifically
RT deaminated in embryonic development.";
RL J. Dermatol. Sci. 37:41-48(2005).
RN [3]
RP IDENTIFICATION, AND TISSUE SPECIFICITY.
RX MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;
RA Hesse M., Zimek A., Weber K., Magin T.M.;
RT "Comprehensive analysis of keratin gene clusters in humans and
RT rodents.";
RL Eur. J. Cell Biol. 83:19-26(2004).
CC -!- TISSUE SPECIFICITY: Expressed in stratified epithelia.
CC -!- PTM: Undergoes deamination of some arginine residues
CC (citrullination).
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and
CC microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II
CC (neutral to basic; 56-70 kDa) [K1 to K8].
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAA01098975; -; NOT_ANNOTATED_CDS; Genomic DNA.
DR EMBL; CAAA01098987; -; NOT_ANNOTATED_CDS; Genomic DNA.
DR EMBL; BK005256; DAA05633.1; -; mRNA.
DR EMBL; BK003993; DAA02234.1; -; mRNA.
DR MGI; MGI:3588209; 4732484G22Rik.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR18993:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Citrullination; Coiled coil; Direct protein sequencing;
KW Intermediate filament; Keratin; Phosphorylation.
FT CHAIN 1 572
FT /FTID=PRO_0000063712.
FT REGION 1 166
FT REGION 167 476 Rod.
FT REGION 167 202 Coil 1A.
FT REGION 203 221 Linker 1.
FT REGION 222 313 Coil 1B.
FT REGION 314 337 Linker 12.
FT REGION 338 476 Coil 2.
FT REGION 477 568 Tail.
FT SITE 420 420 Stutter.
FT MOD_RES 55 55 Phosphoserine (By similarity).
SQ SEQUENCE 572 AA; 61359 MW; 4D6A8BEC8E250096 CRC64;

Query Match 84.4%; Score 38; DB 1; Length 572;
Best Local Similarity 87.5%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEV 9
|||||:|
Db 265 TNAENDFV 272

RESULT 9
Q63ER5_BACCZ
ID Q63ER5_BACCZ PRELIMINARY; PRT; 869 AA.
AC Q63ER5;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Hypothetical protein yhgE.
DE Names=yhgE; OrderedLocustNames=BCE33L0996;
GN Bacillus cereus (strain ZK / E33L).

```



OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
OX Bacillus cereus group.  
OX NCBI\_TaxID=28661;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
RA Hitchcock P., Jackson P., Kelm P., Longmire J., Lucas S., Okinaka R.,  
RA Richardson P., Rubin E., Tice H.;  
RT "Complete genome sequence of Bacillus cereus ZK";  
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CP000001; AAU19250.1; -; Genomic DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 869 AA; 92188 MW; 1B8C5CFD50572342 CRC64;  
  
Query Match 84.4%; Score 38; DB 2; Length 869;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 TNAINEFV 9  
Db 483 TNAQNEFV 490  
|||||  
-----  
RESULT 10  
Q81U13 BACAN  
ID Q81U13 BACAN PRELIMINARY; PRT; 869 AA.  
AC Q81U13; 061298; 06KW31.  
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2003, sequence version 1.  
DT 07-FEB-2006, entry version 18.  
DE Hypothetical protein.  
GN OrderedLocNames=BA1081, BAS1009, GBAA1081, ORFNames=BA\_1081;  
OS Bacillus anthracis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
OX Bacillus cereus group.  
OX NCBI\_TaxID=1392;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Ames / isolate Porton;  
RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,  
RA Nelson K.E., Tetelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,  
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,  
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Raeburn D.,  
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,  
RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,  
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;  
RT "The genome sequence of Bacillus anthracis Ames and comparison to  
RT closely related bacteria";  
RL Nature 423:81-86 (2003).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Ames ancestor;  
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
RA Fraser C.M.;  
RT "Bacillus anthracis comparative genomics";  
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Stearie;  
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
RA Hitchcock P., Jackson P., Kelm P., Longmire J., Lucas S., Okinaka R.,  
RA Richardson P., Rubin E., Tice H.;  
RT "Complete genome sequence of Bacillus anthracis Sterne";

RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AE016879; AAP25060.1; -; Genomic DNA.  
DR EMBL; AE017334; AAT30181.1; -; Genomic DNA.  
DR EMBL; AE017225; AAT53333.1; -; Genomic DNA.  
DR TIGR; BAI081; -;  
DR TIGR; GBAA1081; -;  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 869 AA; 92270 MW; 1466959BF791FA41 CRC64;  
  
Query Match 84.4%; Score 38; DB 2; Length 869;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 TNAINEFV 9  
Db 483 TNAQNEFV 490  
|||||  
-----  
RESULT 11  
Q5H887 NEIGO  
ID Q5H887 NEIGO PRELIMINARY; PRT; 207 AA.  
AC Q5H887;  
DT 01-MAR-2005, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE Fructose-1,6-bisphosphatase (Fragment).  
GN Name=fbp;  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OX Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=485;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=NID103;  
RA Takahashi H., Watanabe H.;  
RT "The pseudogenes highly homologous to Neisseria meningitidis ggt gene  
RT encoding gamma-glutamyl aminopeptidase in the related species:  
RT Implication for the molecular evolution of the meningococcal ggt  
RT gene";  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AB175034; BAD89030.1; -; Genomic DNA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0042578; F:phosphoric ester hydrolase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR000146; In\_Fb\_phptase.  
DR Pfam; PF00316; FBPase; I.  
DR PRINTS; PR00115; FBPHTASE.  
DR PRINTS; PR00377; INFBPHPTASE.  
DR ProDom; PD001491; In\_Fb\_phptase; 1.  
KW Carbohydrate metabolism; Hydrolase.  
FT NON TER 1  
SQ SEQUENCE 207 AA; 23183 MW; F57B82163CA7FC24 CRC64;  
  
Query Match 80.0%; Score 36; DB 2; Length 207;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 3 NAINEFV 9  
Db 58 NAINEFV 64  
|||||  
-----  
RESULT 12  
Q9JUL6 NEIMA  
ID Q9JUL6 NEIMA PRELIMINARY; PRT; 324 AA.

```

AC Q9JUL6;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 21-FEB-2006, entry version 26.
DE Putative fructose-1,6-bisphosphatase (EC 3.1.3.11).
GN Name=fbp; OrderedLocusNames=NMA1259;
OS Neisseria meningitidis serogroup A.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919; DOI=10.1038/35006655;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.M.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K.L., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AL162755; CAB84514.1; -; Genomic_DNA.
DR PIR; A81894; A81894.
DR HSSP; P46275; 1DBZ.
DR BioCyc; NMEN65699:NMA1259-MONOMER; -.
DR GO; GO:0042132; F:fructose-bisphosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0043578; F:phosphoric ester hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000146; In_FBP_phptase.
DR Pfam; PF00316; FBPase; 1.
DR PIRSF; PIRSF000904; FBPTase_SBPase; 1.
DR PRINTS; PR00115; FBPHPTASE.
DR PRINTS; PR00377; INFBPHTASE.
DR ProDom; PD001491; In_FBP_phptase; 1.
KW Carbohydrate metabolism; Complete proteome; Hydrolase.
SQ SEQUENCE 324 AA; 35551 MW; 5F148CBC4AF4B4A9 CRC64;

Query Match 80.0%; Score 36; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9
DB 175 NAENEFV 181

RESULT 13
Q9JZH1 NEIMB Q9JZH1 PRELIMINARY; PRT; 324 AA.
AC Q9JZH1;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 21-FEB-2006, entry version 25.
DE Fructose-1,6-bisphosphatase (EC 3.1.3.11).
GN Name=fbp; OrderedLocusNames=NMB1060;
OS Neisseria meningitidis serogroup B.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=WC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307; DOI=10.1126/science.287.5459.1809;
RA Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C.,
RA Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F.,
RA Dodson R.X., Nelson W.C., Gwinn M.L., DeBoy R.T., Peterson J.D.,

```

```

RA Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
RA Dougherty B.A., Mason T.M., Ciecko A., Parksey D.S., Blair E.,
RA Citron H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
RA Qin H., Vamathevan J.J., Gill J., Scarlato V., Massignani V., Pizzo M.,
RA Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
RA Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE002098; AAF41456.1; -; Genomic_DNA.
DR PIR; F81126; F81126.
DR HSSP; P46275; 1DBZ.
DR TIGR; NMB1060; -.
DR BioCyc; NMEN491:NMB1060-MONOMER; -.
DR GO; GO:0042132; F:fructose-bisphosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0043578; F:phosphoric ester hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000146; In_FBP_phptase.
DR Pfam; PF00316; FBPase; 1.
DR PIRSF; PIRSF000904; FBPTase_SBPase; 1.
DR PRINTS; PR00115; FBPHPTASE.
DR PRINTS; PR00377; INFBPHTASE.
DR ProDom; PD001491; In_FBP_phptase; 1.
KW Carbohydrate metabolism; Complete proteome; Hydrolase.
SQ SEQUENCE 324 AA; 35581 MW; B69F83DCC4D42487 CRC64;

Query Match 80.0%; Score 36; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9
DB 175 NAENEFV 181

RESULT 14
Q5F8C3 NEIG1 Q5F8C3 PRELIMINARY; PRT; 324 AA.
AC Q5F8C3;
DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.
DT 15-MAR-2005, sequence version 1.
DT 07-MAR-2006, entry version 9.
DE Putative fructose-bisphosphatase (EC 3.1.3.11).
GN OrderedLocusNames=NGO0862;
OS Neisseria gonorrhoeae (strain ATCC 700825 / FA 1090).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=24231;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA Lewis L.A., Gillaspay A.F., McLaughlin R.E., Gibson M., Ducey T.F.,
RA Ownbey T., Hartman K., Nydick C., Carlson M.B., Vaughn J., Thomson C.,
RA Song L., Lin S., Yuan X., Najjar F., Zhan M., Ren Q., Zhu H., Qi S.,
RA Kenton S.M., Lai H., White J.D., Clifton S., Roe B.A., Dyer D.W.;
RT "The complete genome sequence of Neisseria gonorrhoeae."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE004969; AAW89564.1; -; Genomic_DNA.
DR GenomeReviews; AE004969.GR; NGO0862.
DR GO; GO:0042132; F:fructose-bisphosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0043578; F:phosphoric ester hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000146; In_FBP_phptase.
DR Pfam; PF00316; FBPase; 1.

```

DR PIRSF: PIRSF000904; FBPTase\_SBPase; 1.  
 DR PRINTS; PR00115; FBPHPTASE.  
 DR PRINTS; PR00377; INFBPHPTASE.  
 DR ProDom; PD001491; In\_FB\_phptase; 1.  
 KW Carbohydrate metabolism; Complete proteome; Hydrolase.  
 SQ SEQUENCE 324 AA; 35474 MW; 013500A67852B411 CRC64;

Query Match 80.0%; Score 36; DB 2; Length 324;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEV 9  
 Db 175 NAENEV 181

RESULT 15  
 Q919P5 RANCA  
 ID Q919P5 RANCA PRELIMINARY; PRT; 529 AA.  
 AC Q919P5  
 DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.  
 DT 01-OCT-2000, sequence version 1.  
 DT 07-FEB-2006, entry version 23.  
 DE Inner-ear cytochrome.  
 OS Rana catesbeiana (Bull frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;  
 OC Aquarana.  
 OX NCBI\_TaxID=8400;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Inner-ear;  
 RX MEDLINE=20243794; PubMed=10758152; DOI=10.1073/pnas.070050797;  
 RA Cyr J.L., Bell A.M., Hudspeth A.J.;  
 RT "Identification with a recombinant antibody of an inner-ear  
 cytochrome, a marker for hair-cell differentiation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:4908-4913(2000).  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AF229168; AAF34720.1; -; mRNA.  
 DR HSSP; P08670; ICK7.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin I.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR18893.SFS; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament.  
 SQ SEQUENCE 529 AA; 56619 MW; 3B6125D4AAEC108E CRC64;

Query Match 80.0%; Score 36; DB 2; Length 529;  
 Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TSAENEV 9  
 Db 230 TSAENEV 237

Search completed: July 5, 2006, 19:01:30  
 Job time : 93.1971 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 78.2039 Seconds

(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928A-8

Perfect score: 45

Sequence: 1 ATNAENEFV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 8:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

9: Geneseqp2005s:\*

10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	9	6 AAO26388	AAO26388 Psoriasis
2	45	100.0	9	9 ADV04413	Adv04413 Leishmani
3	41	91.1	11	6 ABG74868	Abg74868 Human cyt
4	41	91.1	113	6 ABG74871	Abg74871 Human cyt
5	41	91.1	628	8 ADS85115	Ads85115 Mouse ato
6	41	91.1	643	6 ABG74869	Abg74869 Human cyt
7	41	91.1	644	8 ADQ17549	Adq17549 Human sof
8	41	91.1	644	9 AEA15501	Aea15501 Human pol
9	36	80.0	11	4 AAU28642	Aau28642 DPI trypt
10	36	80.0	11	4 AAU26290	Aau26290 Depressio
11	36	80.0	324	8 ADJ49030	Adj49030 Oil-assoc
12	36	80.0	324	8 ADJ50324	Adj50324 Oil-assoc
13	36	80.0	395	6 ABP79368	Abp79368 N. gonorr
14	36	80.0	583	6 ABU26459	Abu26459 Protein e
15	36	80.0	600	9 AEA15488	Aea15488 Human pol
16	35	77.8	11	4 AAU68144	Aau68144 Human Bre
17	35	77.8	11	4 AAU68118	Aau68118 Human Bre
18	35	77.8	11	4 AAU68063	Aau68063 Human Bre
19	35	77.8	11	4 AAU68448	Aau68448 Human Bre
20	35	77.8	20	9 AED42280	Aed42280 Human pro
21	35	77.8	342	7 ADE15641	Ade15641 Human str
22	35	77.8	508	8 ADS10914	Ads10914 Human the
23	35	77.8	521	5 ABG94649	Abg94649 Human NOV

24	35	77.8	521	7 ADL36002	Adl36002 Human NOV
25	35	77.8	526	5 ABJ04655	Abj04655 Protein o
26	35	77.8	529	5 ABB83482	Abb83482 Human cyt
27	35	77.8	529	9 AEA15497	Aea15497 Human pol
28	35	77.8	536	5 ABJ04646	Abj04646 Protein o
29	35	77.8	540	9 AEA15499	Aea15499 Human pol
30	35	77.8	543	9 AEA20229	Aea20229 Novel hum
31	35	77.8	567	7 ADB5618	Adb5618 Human pro
32	35	77.8	590	7 ADB70328	Adb70328 Keratin 5
33	35	77.8	590	7 ADB75394	Adb75394 Prostate
34	35	77.8	590	8 ADJ37107	Adj37107 Human mal
35	35	77.8	590	8 ADR99136	Adr99136 Keratin 5
36	35	77.8	590	9 AEA15495	Aea15495 Human pol
37	35	77.8	590	9 AEB87738	Aeb87738 Human cyt
38	35	77.8	601	8 ADM07163	Adm07163 Bos tauru
39	35	77.8	602	6 ABR41632	Abr41632 Human DIT
40	34	75.6	271	5 AAU96118	Aau96118 Ehrlichia
41	34	75.6	341	5 ABB54677	Abb54677 Lactococc
42	34	75.6	341	8 ADS29399	Ads29399 Bacterial
43	34	75.6	659	6 ABU35613	Abu35613 Protein e
44	34	75.6	872	6 ABP76774	Abp76774 N. gonorr
45	34	75.6	886	6 ABP79007	Abp79007 N. gonorr

## ALIGNMENTS

### RESULT 1

AAO26388

ID AAO26388 standard; peptide; 9 AA.

AC AAO26388;

XX

DT 30-JAN-2003 (first entry)

XX

DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 8.

XX

KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.

XX

OS Leishmania sp.

XX

PN WO200274239-A2.

XX

PD 26-SEP-2002.

XX

PF 04-MAR-2002; 2002WO-US006496.

XX

PR 16-MAR-2001; 2001US-00809003.

XX

PA (AKIV-) AKIVA LLC.

XX

PI O'daly JA;

XX

DR WPI; 2003-018763/01.

XX

PT New polypeptide and immunogenic variants comprising amino acid sequences  
of particulate antigens, useful for the treatment and clinical remission  
of psoriasis.

XX

PS Claim 1; Page 43; 56pp; English.

XX

CC The invention relates to a polypeptide comprising an isolated amino acid  
sequence or immunogenic variants selected from any of 14 fully defined  
sequences of 7-16 amino acids, given in the specification. The

CC immunotherapeutic agents and a pharmaceutical compositions comprising  
polynucleotides and vectors of the invention are useful for the treatment  
and clinical remission of psoriasis. The isolated nucleic acids are

CC useful as probes. The sequences of the invention can be used in the  
treatment of disorders by gene therapy. This sequence represents one of  
the 14 immunotherapeutic peptides of the invention

XX

## SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 6; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATNANEFV 9  
 |||||  
 Db 1 ATNANEFV 9

## RESULT 2

ADV04413  
 ID ADV04413 standard; peptide; 9 AA.

XX  
 AC ADV04413;

XX 24-FEB-2005 (first entry)

DE Leishmania immunotherapeutic peptide SEQ ID NO:8.

XX T-lymphocyte; immunogenicity; antigen; antiprosoritic; immunostimulant;  
 KW psoriasis.

XX Leishmania sp.

XX US2004241168-A1.

XX 02-DEC-2004.

XX 09-FEB-2004; 2004US-00774928.

XX 16-MAR-2001; 2001US-00809003.

XX 17-OCT-2003; 2003US-00687892.

XX (ODAL/) ODALY J A.

XX Odaly JA;

XX WPI; 2005-011563/01.

XX Inhibiting selectively T-cell rolling in human, by administering compound  
 PT interfering with cutaneous lymphocyte antigen-E selectin interaction and  
 PT leukocyte function associated antigen-1 interaction.

PS Claim 18; SEQ ID NO 8; 21pp; English.

XX The invention relates to a novel method for selectively inhibiting T-cell  
 CC rolling in a human host, comprising administering a compound that  
 CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
 CC selectin interaction and leukocyte function associated antigen (LFA)-  
 CC 1/intercellular adhesion molecule (ICAM) and very late antigen  
 CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
 CC includes an immunotherapeutic agent, which comprises a purified protein  
 CC extract that is isolated by diethylaminoethyl Sephadex chromatography of  
 CC a Nonidet P-40 insoluble particulate antigen fraction derived from  
 CC isolated killed cells of amastigotes from one or more species of the  
 CC Leishmania genus, where the particulate antigen fraction is solubilized  
 CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
 CC to diethylaminoethyl Sephadex and is eluted with a solution comprising  
 CC 0.1 M sodium chloride, 8 M urea and 0.025 M  
 CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract  
 CC includes polypeptides having apparent molecular weights of 73, 80 and 82  
 CC kDa, after total reduction and alkylation. The species is L. amazonensis,  
 CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the  
 CC invention has antiprosoritic and immunostimulant activity, and inhibits T-  
 CC cell rolling by interfering with CLA-E selectin interaction and LFA-  
 CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
 CC inhibiting T-cell rolling in a human host, and for treating psoriasis.  
 CC The present sequence represents a Leishmania peptide used in the  
 CC invention.

XX Sequence 9 AA;

Query Match 100.0%; Score 45; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATNANEFV 9  
 |||||  
 Db 1 ATNANEFV 9

## RESULT 3

ABG74868  
 ID ABG74868 standard; peptide; 11 AA.

XX  
 AC ABG74868;

XX 19-JUN-2003 (first entry)

DE Human cytokerin-1 associated peptide SEQ ID 2.

XX Cytokeratin-1; inflammation; infection; soluble; sepsis; veterinary;  
 KW sepsis-like systemic infection; human.

XX Homo sapiens.

XX WO2003002600-A1.

XX 09-JAN-2003.

XX 12-JUN-2002; 2002WO-EP006473.

XX 27-JUN-2001; 2001DE-01030985.

XX (BRAH-) BRAHMS AG.

XX Bergmann A, Struck J, Uehlein M;

XX WPI; 2003-201491/19.

XX New use of soluble cytokerin-1 fragments from body fluid are tissue as  
 PT marker peptides or to provide therapeutics for the diagnosis, prognosis  
 PT and treatment of inflammatory disease and infection, including systemic  
 PT inflammation.

PS Disclosure; Page 38; 51pp; German.

XX This invention describes a novel use of soluble cytokerin-1 fragments  
 CC from body fluid or tissue as marker peptides for the detection, prognosis  
 CC and control of inflammation and infection or as a therapeutic target for  
 CC inflammation and infection. The invention also describes a method of  
 CC producing a soluble cytokerin-1 fragment in rich or pure form from a  
 CC human material rich in cytokerin-1 by enzymatically splitting the  
 CC cytokerin-1 with endoprotease to obtain the hydrolysis products and  
 CC purifying the resulting fragments. The novel cytokerin fragments are  
 CC useful for the differential diagnosis, early recognition and prognosis,  
 CC evaluation of severity of sepsis and heavy infection, particularly sepsis,  
 CC -like systemic infection. The fragments are used in human and veterinary  
 CC therapy, particularly the fragment or it's specific antibody is used to  
 CC treat inflammatory disease or infection. This sequence represents a  
 CC cytokerin-1 associated peptide described in the disclosure of the  
 CC invention

XX Sequence 11 AA;

Query Match 91.1%; Score 41; DB 6; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9  
 |||||  
 Db 1 TNAENEFV 8



```
XX
PR 27-JUN-2001; 2001DE-01030985.
XX
PA (BRAH-) BRAHMS AG.
XX
PI Bergmann A, Struck J, Uehlein M;
XX
XX WPI; 2003-201491/19.
XX
PT New use of soluble cytokeratin-1 fragments from body fluid are tissue as
PT marker peptides or to provide therapeutics for the diagnosis, prognosis
PT and treatment of inflammatory disease and infection, including systemic
PT inflammation.
XX
XX Claim 3; Page 38-40; 51pp; German.
XX
XX This invention describes a novel use of soluble cytokeratin-1 fragments
XX from body fluid or tissue as marker peptides for the detection, prognosis
XX and control of inflammation and infection or as a therapeutic target for
XX inflammation and infection. The invention also describes a method of
XX producing a soluble cytokeratin-1 fragment in rich or pure form from a
XX human material rich in cytokeratin-1 by enzymatically splitting the
XX cytokeratin-1 with endoprotease to obtain the hydrolysis products and
XX purifying the resulting fragments. The novel cytokeratin fragments are
XX useful for the differential diagnosis, early recognition and prognosis,
XX evaluation of severity of sepsis and heavy infection, particularly sepsis
XX -like systemic infection. The fragments are used in human and veterinary
XX therapy, particularly the fragment or it's specific antibody is used to
XX treat inflammatory disease or infection. This sequence represents the
XX human cytokeratin-1 protein described in the disclosure of the invention
XX
XX Sequence 643 AA;
SQ
Query Match 91.1%; Score 41; DB 6; Length 643;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 TNAENEV 9
Db 277 TNAENEV 284
|||||||
RESULT 7
ID ADQ17549 standard; protein; 644 AA.
XX
XX ADQ17549;
AC
XX
XX 26-AUG-2004 (first entry)
DT
XX
XX Human soft tissue sarcoma-upregulated protein - SEQ ID 366.
DE
XX
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
KW
XX
XX Homo sapiens.
OS
XX
XX WO2004048938-A2.
PN
XX
XX 10-JUN-2004.
PD
XX
XX 26-NOV-2003; 2003WO-US038193.
PF
XX
XX 26-NOV-2002; 2002US-0429739P.
PR
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
PA
XX
XX Aziz N, Ginsburg WM, Zlotnik A;
PI
XX
XX WPI; 2004-441208/41.
DR
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT
```

```
PT sarcoma.
XX
XX Example 2; SEQ ID NO 366; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytostatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX protein of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 644 AA;
SQ
Query Match 91.1%; Score 41; DB 8; Length 644;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 TNAENEV 9
Db 278 TNAENEV 285
|||||||
RESULT 8
ID AEA15501 standard; protein; 644 AA.
XX
XX AEA15501;
AC
XX
XX 28-JUL-2005 (first entry)
DT
XX
XX Human polypeptide #120.
DE
XX
XX Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;
KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;
KW cytostatic; neoplasm.
XX
XX Homo sapiens.
OS
XX
XX WO2005047534-A2.
PN
XX
XX 26-MAY-2005.
PD
XX
XX 15-OCT-2004; 2004WO-EP011599.
PF
XX
XX 28-OCT-2003; 2003EP-00024565.
PR
XX
XX (FARB ) BAYER HEALTHCARE AG.
PA
XX
XX Wirtz R, Munnes M;
PI
XX
XX WPI; 2005-372393/38.
DR
XX
XX N-PSDB; AEA15423.
DR
XX
XX Predicting a response to cancer treatment by detecting at least 2
PT markers, which are genes or genomic nucleic acid sequences that are
PT located on one chromosomal region, which is altered in malignant
PT neoplasia.
XX
XX Claim 7; SEQ ID NO 463; 464pp; English.
XX
XX The invention relates to a method of predicting response to cancer
XX treatment comprising detection of at least 2 markers, where the markers
XX are genes and fragments or genomic nucleic acid sequences that are
XX located on one chromosomal region, which is altered in malignant
XX neoplasia. The invention also relates to a method for the prediction,
XX diagnosis or prognosis of malignant neoplasia, methods for detecting
XX
```



CC deregulations in malignant neoplasia and breast cancer, a method of  
 CC determining the phenotype of a cell or tissue, a method for identifying  
 CC genomic regions which are altered on the chromosomal level and encode  
 CC genes that are linked by function and are differentially expressed in  
 CC malignant neoplasia and breast cancer, methods of screening for agents  
 CC which regulate the activity of a polypeptide or a polynucleotide and  
 CC antibodies that specifically bind to a full length or partial  
 CC polypeptide. The method is useful for predicting response to cancer  
 CC treatment. The methods and compositions are useful for predicting,  
 CC diagnosing, prognosing, preventing or treating malignant neoplasia  
 CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
 CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
 CC lung cancer. This sequence represents a human polypeptide used in the  
 CC scope of the invention.

XX Sequence 644 AA;

Query Match 91.1%; Score 41; DB 9; Length 644;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9  
 |||||  
 Db 278 TNAENEFV 285

RESULT 9  
 AAU28642  
 ID AAU28642 standard; peptide; 11 AA.

XX

AC AAU28642;

XX

DT 03-JAN-2002 (first entry)

XX

DE DPI tryptic digest peptide #239.

XX

XX Human; depression associated protein isoform; tryptic digest peptide;

XX DPI; cerebrospinal fluid; CSF; BAD; bipolar affective disorder;

XX neuropsychiatric disorder; bipolar mood disorder; neuroleptic;

XX maniac-depressive illness; schizoaffective disorder.

XX Homo sapiens.

XX

XX WO200162787-A1.

XX

XX 30-AUG-2001.

XX

XX 23-FEB-2001; 2001WO-GB000786.

XX

XX 24-FEB-2000; 2000GB-00004412.

XX

XX 08-DEC-2000; 2000GB-00030050.

XX

XX 12-DEC-2000; 2000US-0254830P.

XX

XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX

XX Herath HMAC, Parekh RB, Rohlf C, Terrett JA, Tyson KL;

CC screening, diagnosis, prognosis, therapy and prophylaxis of  
 CC neuropsychiatric disorders e.g. BAD (also known as bipolar mood disorder,  
 CC BP), maniac-depressive illnesses, attention deficit disorders,  
 CC schizoaffective disorders, and unipolar affective disorders. The present  
 CC sequence represents one of the DPI tryptic digest peptides of the present  
 CC invention  
 XX

SQ Sequence 11 AA;

Query Match 80.0%; Score 36; DB 4; Length 11;  
 Best Local Similarity 87.5%; Pred. No. 1.9;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9  
 |||||  
 Db 1 TDAENEFV 8

RESULT 10

AAU26290

ID AAU26290 standard; peptide; 11 AA.

XX

AC AAU26290;

XX

DT 18-DEC-2001 (first entry)

XX

DE Depression-Associated Protein isoform DPI-60.

XX

XX Human; Bipolar Affective Disorder; BAD; Depression-Associated feature;

XX DF; Depression-Associated protein isoform; DPI; Cerebro-spinal fluid;

XX CSF; antidepressant; antimanic; nootropic; tranquiliser; neuroleptic;

XX attention deficient disorder; schizoaffective disorder;

XX unipolar affective disorder.

XX Homo sapiens.

XX WO200163294-A2.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-GB000791.

XX 24-FEB-2000; 2000GB-00004412.

XX 08-DEC-2000; 2000GB-00030050.

XX 12-DEC-2000; 2000US-0254830P.

XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX Herath HMAC, Parekh RB, Rohlf C;

XX WPI; 2001-592081/65.

XX Preparation for diagnosing or treating bipolar affected disorder (BAD) or

XX unipolar depression, or for screening for modulators, comprises a BAD-

XX associated protein isoform.

XX Claim 8; Page 36; 163pp; English.

XX The invention relates to a preparation comprising an isolated Bipolar

XX Affected Disorder (BAD)-Associated Protein Isoform (DPIs). The DPI's are

XX used to screen, diagnose or prognosis of BAD or unipolar depression,

XX determine the stage or severity of BAD or unipolar depression, identify a

XX subject at risk of developing BAD or unipolar depression, or monitor the

XX effect of therapy in a subject. They are also used to screen for or

XX identify agents that interact with a DPI. These agents, antibodies  
 CC or prevent BAD or unipolar depression. Diseases that can be treated are  
 CC attention deficient disorder, a schizoaffective disorder, a bipolar or a  
 CC unipolar affective disorder. The DPIs are used in proteomics. The  
 CC proteomic approach of using DPIs for screening, diagnosis or prognosis of  
 CC BAD or unipolar depression overcomes the problems of using gene  
 CC expression analysis, such as not being able to obtain central nervous  
 CC system (CNS) tissue from a living patient under normal circumstances. The

XX WPI; 2001-570626/64.

XX Novel nucleic acid encoding a protein associated with bipolar affective  
 PT disorder, which is used for diagnosis, prophylaxis and therapy of  
 PT neuropsychiatric disorders, such as bipolar affective disorder.

XX Disclosure; Page 36; 153pp; English.

XX The present invention relates to the identification of depression  
 CC associated protein isoforms (DPIs), particularly the tryptic digest  
 CC peptides of these proteins. Some of the DPIs (AAU28404-AAU28625)  
 CC described are decreased in the cerebrospinal fluid (CSF) of BAD (bipolar  
 CC affective disorder) subjects, whilst other DPIs (AAU28626-AAU28887) are  
 CC increased in BAD subjects. Also described are peptide sequences  
 CC identified from DPI-45 and DPI-213 and the nucleic acid sequence they are  
 CC encoded by. The sequences of the invention are useful for clinical

CC present sequence is a DIP increased in the CSF (cerebro-spinal fluid) of  
CC subjects having BAD

XX Sequence 11 AA;

Query Match 80.0%; Score 36; DB 4; Length 11;  
Best Local Similarity 87.5%; Pred. No. 1.9;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEFV 9  
|:|||||  
Db 1 TDAENEFV 8

## RESULT 11

ADJ49030  
ID ADJ49030 standard; protein; 324 AA.

XX AC ADJ49030;  
XX DT 06-MAY-2004 (first entry)  
XX DE Oil-associated gene related protein #530.  
XX KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.  
XX OS Unidentified.

XX PN US2004025202-A1.  
XX PD 05-FEB-2004.  
XX PF 14-MAR-2003; 2003US-00389566.

XX PR 15-MAR-2002; 2002US-0365301P.  
XX PR 26-JUN-2002; 2002US-0391786P.  
XX PR 26-JUN-2002; 2002US-0392018P.

XX PA (LAUR/) LAURIE C C.  
XX PA (RAVA/) RAVANELLO M.  
XX PA (SAVA/) SAVAGE T.  
XX PA (LEDE/) LEDEAUX J R.  
XX PA (ROGE/) ROGERS J A.

XX PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;  
XX WPI; 2004-142683/14.

XX PT Novel recombinant DNA construct comprising a promoter functional in  
XX plants operably linked to an oil-associated gene for producing transgenic  
XX plant seed.

XX Example 3; SEQ ID NO 1034; 22pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
XX promoter functional in plants operably linked to an oil-associated gene.  
XX The construct is useful for transgenic plant seed which has in its genome  
XX the construct, that is functional in the plant to transcribe the oil-  
XX associated gene. The transgenic plant seed grows into a plant having  
XX enhanced seed oil as compared to wild type. The construct is useful for  
XX producing hybrid maize seed. The transgenic plant seed is useful for  
XX producing vegetable oil. The present sequence represents the amino acid  
XX sequence of an oil-associated gene related protein.

XX Sequence 324 AA;

Query Match 80.0%; Score 36; DB 8; Length 324;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9  
|||||||  
Db 175 NAENEFV 181

## RESULT 12

ADJ50324  
ID ADJ50324 standard; protein; 324 AA.

XX AC ADJ50324;  
XX DT 06-MAY-2004 (first entry)  
XX DE Oil-associated gene related protein #1824.  
XX KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.  
XX OS Unidentified.

XX PN US2004025202-A1.  
XX PD 05-FEB-2004.  
XX PF 14-MAR-2003; 2003US-00389566.

XX PR 15-MAR-2002; 2002US-0365301P.  
XX PR 26-JUN-2002; 2002US-0391786P.  
XX PR 26-JUN-2002; 2002US-0392018P.

XX PA (LAUR/) LAURIE C C.  
XX PA (RAVA/) RAVANELLO M.  
XX PA (SAVA/) SAVAGE T.  
XX PA (LEDE/) LEDEAUX J R.  
XX PA (ROGE/) ROGERS J A.

XX PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;  
XX WPI; 2004-142683/14.

XX PT Novel recombinant DNA construct comprising a promoter functional in  
XX plants operably linked to an oil-associated gene for producing transgenic  
XX plant seed.

XX Example 3; SEQ ID NO 2328; 22pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
XX promoter functional in plants operably linked to an oil-associated gene.  
XX The construct is useful for transgenic plant seed which has in its genome  
XX the construct, that is functional in the plant to transcribe the oil-  
XX associated gene. The transgenic plant seed grows into a plant having  
XX enhanced seed oil as compared to wild type. The construct is useful for  
XX producing hybrid maize seed. The transgenic plant seed is useful for  
XX producing vegetable oil. The present sequence represents the amino acid  
XX sequence of an oil-associated gene related protein.

XX Sequence 324 AA;

Query Match 80.0%; Score 36; DB 8; Length 324;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEFV 9  
|||||||  
Db 175 NAENEFV 181

## RESULT 13

ABP79368  
ID ABP79368 standard; protein; 395 AA.

XX AC ABP79368;  
XX DT 07-MAR-2003 (first entry)

XX DE N. gonorrhoeae amino acid sequence SEQ ID 5266.

XX Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB002069.  
XX PR 12-FEB-2001; 2001GB-00003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Fontana MR, Pizza M, Massignani V, Monaci E;  
XX DR WPI; 2003-058415/05.  
XX DR N-PSDB; ABZ40338.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX PS medicament for treating or preventing N. gonorrhoeae infection.  
XX PS Disclosure; Page 574; 815pp; English.  
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.  
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and  
XX CC antibodies that specifically bind to the proteins. The composition  
XX CC comprising the protein, nucleic acid or antibody is useful for the  
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX CC infection, this may be in the form of a vaccine or gene therapy.  
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX CC molecules of the invention  
XX SQ Sequence 395 AA;  
Query Match 80.0%; Score 36; DB 6; Length 395;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 NAENEV 9  
Db 246 NAENEV 252  
|||||  
RESULT 14  
ABU26459  
ID ABU26459 standard; protein; 583 AA.  
XX AC ABU26459;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #11986.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Campylobacter jejuni.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB002069.  
XX PR 12-FEB-2001; 2001GB-00003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Fontana MR, Pizza M, Massignani V, Monaci E;  
XX DR WPI; 2003-058415/05.  
XX DR N-PSDB; ABZ40338.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX PS medicament for treating or preventing N. gonorrhoeae infection.  
XX PS Disclosure; Page 574; 815pp; English.  
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.  
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and  
XX CC antibodies that specifically bind to the proteins. The composition  
XX CC comprising the protein, nucleic acid or antibody is useful for the  
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX CC infection, this may be in the form of a vaccine or gene therapy.  
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX CC molecules of the invention  
XX SQ Sequence 395 AA;  
Query Match 80.0%; Score 36; DB 6; Length 395;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 NAENEV 9  
Db 246 NAENEV 252  
|||||  
RESULT 14  
ABU26459  
ID ABU26459 standard; protein; 583 AA.  
XX AC ABU26459;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #11986.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Campylobacter jejuni.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB002069.  
XX PR 12-FEB-2001; 2001GB-00003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Fontana MR, Pizza M, Massignani V, Monaci E;  
XX DR WPI; 2003-058415/05.  
XX DR N-PSDB; ABZ40338.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX PS medicament for treating or preventing N. gonorrhoeae infection.  
XX PS Disclosure; Page 574; 815pp; English.  
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.  
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and  
XX CC antibodies that specifically bind to the proteins. The composition  
XX CC comprising the protein, nucleic acid or antibody is useful for the  
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX CC infection, this may be in the form of a vaccine or gene therapy.  
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX CC molecules of the invention  
XX SQ Sequence 395 AA;  
Query Match 80.0%; Score 36; DB 6; Length 583;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATNAENEV 9  
Db 61 ASARNEV 69  
|:|||||  
RESULT 15  
AEA15488  
ID AEA15488 standard; protein; 600 AA.  
XX AC AEA15488;  
XX DT 28-JUL-2005 (first entry)  
XX DE Human polypeptide #107.  
XX KW Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
XX KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
XX KW cytostatic; neoplasm.  
XX OS Homo sapiens.  
XX PN WO2005047534-A2.  
XX PD 26-MAY-2005.

```

XX 15-OCT-2004; 2004WO-EP011599.
PF
XX
XX 28-OCT-2003; 2003EP-00024565.
PR
XX
XX (FARB ) BAYER HEALTHCARE AG.
PA
XX
XX Wirtz R, Munnes M;
PI
XX
XX WPI; 2005-372393/38.
DR
XX N-PSDB; AEA15410.
DR
XX
XX Predicting a response to cancer treatment by detecting at least 2
PT markers, which are genes or genomic nucleic acid sequences that are
PT located on one chromosomal region, which is altered in malignant
PT neoplasia.
XX
XX Claim 7; SEQ ID NO 450; 464pp; English.
PS
XX
XX The invention relates to a method of predicting response to cancer
CC treatment comprising detection of at least 2 markers, where the markers
CC are genes and fragments or genomic nucleic acid sequences that are
CC located on one chromosomal region, which is altered in malignant
CC neoplasia. The invention also relates to a method for the prediction,
CC diagnosis or prognosis of malignant neoplasia, methods for detecting
CC deregulations in malignant neoplasia and breast cancer, a method of
CC determining the phenotype of a cell or tissue, a method for identifying
CC genomic regions which are altered on the chromosomal level and encode
CC genes that are linked by function and are differentially expressed in
CC malignant neoplasia and breast cancer, methods of screening for agents
CC which regulate the activity of a polypeptide or a polynucleotide and
CC antibodies that specifically bind to a full length or partial
CC polypeptide. The method is useful for predicting response to cancer
CC treatment. The methods and compositions are useful for predicting,
CC diagnosing, prognosing, preventing or treating malignant neoplasia
CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,
CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell
CC lung cancer. This sequence represents a human polypeptide used in the
CC scope of the invention.
XX
SQ Sequence 600 AA;
Query Match 80.0%; Score 36; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 NAENEV 9
Db 262 NAENEV 268
|||||

```

Search completed: July 5, 2006, 18:43:51  
Job time : 80.2039 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 65.8835 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-8

Perfect score: 45

Sequence: 1 ATNAENEFV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main.\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	9	5	US-10-774-928-8
2	41	91.1	11	5	US-10-480-848A-2
3	41	91.1	11	5	US-10-480-848A-5
4	41	91.1	643	5	US-10-480-848A-3
5	41	91.1	644	5	US-10-723-860-366
6	36	80.0	11	3	US-09-791-393-170
7	36	80.0	11	3	US-09-791-389-170
8	36	80.0	324	4	US-10-389-566-1034
9	36	80.0	324	4	US-10-389-566-2328
10	36	80.0	324	5	US-10-732-923-17878
11	36	80.0	324	5	US-10-732-923-17879
12	36	80.0	324	6	US-11-188-298-3960
13	36	80.0	324	6	US-11-188-298-15864
14	36	80.0	395	5	US-10-467-657-5266
15	36	80.0	583	4	US-10-282-122A-54383
16	35	77.8	11	3	US-09-988-493-41
17	35	77.8	20	6	US-11-075-234-346
18	35	77.8	222	4	US-10-231-913-263
19	35	77.8	315	4	US-10-231-913-259
20	35	77.8	318	4	US-10-231-913-261
21	35	77.8	521	4	US-10-015-115-48
22	35	77.8	526	4	US-10-231-913-32
23	35	77.8	529	4	US-10-250-613-11
24	35	77.8	529	6	US-11-124-368A-182
25	35	77.8	536	4	US-10-231-913-14
26	35	77.8	567	4	US-10-104-047-3772
27	35	77.8	567	6	US-11-072-512-3772

28 35 77.8 590 4 US-10-205-823-218 Sequence 218, Appl  
29 35 77.8 590 4 US-10-236-031B-20 Sequence 20, Appl  
30 35 77.8 590 4 US-10-788-792-142 Sequence 142, Appl  
31 35 77.8 590 6 US-11-051-454-218 Sequence 218, Appl  
32 35 77.8 590 6 US-11-037-713-8 Sequence 8, Appl  
33 34 75.6 271 3 US-09-811-007-46 Sequence 46, Appl  
34 34 75.6 271 4 US-10-062-624-46 Sequence 46, Appl  
35 34 75.6 271 4 US-10-062-051-46 Sequence 46, Appl  
36 34 75.6 271 4 US-10-062-920-46 Sequence 46, Appl  
37 34 75.6 271 4 US-10-680-349-46 Sequence 46, Appl  
38 34 75.6 271 5 US-10-731-554-46 Sequence 46, Appl  
39 34 75.6 341 4 US-10-369-493-18432 Sequence 18432, A  
40 34 75.6 659 4 US-10-282-122A-63537 Sequence 63537, A  
41 34 75.6 872 5 US-10-467-657-78 Sequence 78, Appl  
42 34 75.6 886 5 US-10-467-657-4544 Sequence 4544, Ap  
43 34 75.6 1336 4 US-10-437-963-199521 Sequence 199521,  
44 33 73.3 11 3 US-09-988-493-14 Sequence 14, Appl  
45 33 73.3 11 3 US-09-988-493-66 Sequence 66, Appl

#### ALIGNMENTS

##### RESULT 1

US-10-774-928-8

; Sequence 8, Application US/10774928

; Publication No. US20040241168A1

; GENERAL INFORMATION:

; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission

; TITLE OF INVENTION: Psoriasis

; FILE REFERENCE: 302178

; CURRENT APPLICATION NUMBER: US/10/774,928

; CURRENT FILING DATE: 2004-02-09

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 8

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Leishmania

US-10-774-928-8

Query Match 100.0%; Score 45; DB 5; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.9e+06;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATNAENEFV 9

Db 1 ATNAENEFV 9

##### RESULT 2

US-10-480-848A-2

; Sequence 2, Application US/10480848A

; Publication No. US20040219597A1

; GENERAL INFORMATION:

; APPLICANT: B.R.A.H.M.S Aktiengesellschaft

; TITLE OF INVENTION: Use of soluble cytochrome 1 fragments in diagnosis and

; FILE REFERENCE: 3537PCTNAT

; CURRENT APPLICATION NUMBER: US/10/480,848A

; CURRENT FILING DATE: 2003-12-15

; PRIOR APPLICATION NUMBER: DE 101 30 985.6

; PRIOR FILING DATE: 2001-06-27

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 11

; TYPE: PRT

; ORGANISM: Primat (Pavian)

US-10-480-848A-2

Query Match 91.1%; Score 41; DB 5; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNANEFV 9  
| | | | |  
Db 1 TNANEFV 8

## RESULT 3

US-10-480-848A-5  
; Sequence 5, Application US/10480848A  
; Publication No. US20040219597A1  
; GENERAL INFORMATION:  
; APPLICANT: B.R.A.H.M.S Aktiengesellschaft  
; TITLE OF INVENTION: Use of soluble cytokeatin 1 fragments in diagnosis and  
; TITLE OF INVENTION: therapy  
; FILE REFERENCE: 3537PCTWAT  
; CURRENT APPLICATION NUMBER: US/10/480,848A  
; CURRENT FILING DATE: 2003-12-15  
; PRIOR APPLICATION NUMBER: DE 101 30 985.6  
; PRIOR FILING DATE: 2001-06-27  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-480-848A-5

Query Match 91.1%; Score 41; DB 5; Length 113;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNANEFV 9  
| | | | |  
Db 93 TNANEFV 100

## RESULT 4

US-10-480-848A-3  
; Sequence 3, Application US/10480848A  
; Publication No. US20040219597A1  
; GENERAL INFORMATION:  
; APPLICANT: B.R.A.H.M.S Aktiengesellschaft  
; TITLE OF INVENTION: Use of soluble cytokeatin 1 fragments in diagnosis and  
; TITLE OF INVENTION: therapy  
; FILE REFERENCE: 3537PCTWAT  
; CURRENT APPLICATION NUMBER: US/10/480,848A  
; CURRENT FILING DATE: 2003-12-15  
; PRIOR APPLICATION NUMBER: DE 101 30 985.6  
; PRIOR FILING DATE: 2001-06-27  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; TITLE: L.Johnson et al., Structure of a gene for the human  
; TITLE: epidermal 67-kDa keratin  
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
; VOLUME: 1985  
; ISSUE: 82  
; PAGES: 1896-1900  
US-10-480-848A-3

Query Match 91.1%; Score 41; DB 5; Length 643;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNANEFV 9  
| | | | |  
Db 277 TNANEFV 284

## RESULT 5

US-10-723-860-366  
; Sequence 366, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Naraeba  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnick, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 366  
; LENGTH: 644  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-723-860-366

Query Match 91.1%; Score 41; DB 5; Length 644;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNANEFV 9  
| | | | |  
Db 278 TNANEFV 285

## RESULT 6

US-09-791-393-170  
; Sequence 170, Application US/09791393  
; Publication No. US20030032200A1  
; GENERAL INFORMATION:  
; APPLICANT: Herath, Herath Mudiyanseelage Athula Chandrasiri  
; APPLICANT: Parekh, Rajesh Bhikhu  
; APPLICANT: Rohlf, Christian  
; TITLE OF INVENTION: Proteins, Genes and Their Use for  
; TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)  
; TITLE OF INVENTION: and Unipolar Depression  
; FILE REFERENCE: 2543-1-001 N1  
; CURRENT APPLICATION NUMBER: US/09/791,393  
; CURRENT FILING DATE: 2002-01-02  
; EARLIER APPLICATION NUMBER: GB 0004412.3  
; EARLIER FILING DATE: 2000-02-24  
; EARLIER APPLICATION NUMBER: GB 0030050.9  
; EARLIER FILING DATE: 2000-12-08  
; EARLIER APPLICATION NUMBER: US 60/254,830  
; EARLIER FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 308  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 170  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: homo sapien  
US-09-791-393-170

Query Match 80.0%; Score 36; DB 3; Length 11;  
Best Local Similarity 87.5%; Pred. No. 3.6;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNANEFV 9  
| | | | |  
Db 1 TDANEFV 8

## RESULT 7

US-09-791-389-170

; Sequence 170, Application US/097911389  
; Publication No. US20030032773A1  
; GENERAL INFORMATION:  
; APPLICANT: Herath, Herath Mudiyansele Athula Chandrasiri  
; APPLICANT: Parekh, Rajesh Bhikhu  
; APPLICANT: Rohlf, Christian  
; APPLICANT: Terrett, Jonathan Alexander  
; APPLICANT: Tyson, Kerry Louise  
; TITLE OF INVENTION: Proteins, Genes and Their Use for  
; TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)  
; TITLE OF INVENTION: and Unipolar Depression  
; FILE REFERENCE: 2543-1-001 N2  
; CURRENT APPLICATION NUMBER: US/09/791,389  
; CURRENT FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: GB 0004412.3  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: GB 0030050.9  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: US 60/254,830  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 308  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 170  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: homo sapien  
US-09-791-389-170

Query Match 80.0%; Score 36; DB 3; Length 11;  
Best Local Similarity 87.5%; Pred. No. 3.6;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TNAENFV 9  
Db 1 TDAENFV 8

RESULT 8  
US-10-389-566-1034  
; Sequence 1034, Application US/10389566  
; Publication No. US20040025202A1  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC  
; APPLICANT: Laurie, Cathy C  
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants  
; FILE REFERENCE: 38-77(52900)D  
; CURRENT APPLICATION NUMBER: US/10/389,566  
; CURRENT FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US 60/365,301  
; PRIOR FILING DATE: 2002-03-15  
; PRIOR APPLICATION NUMBER: US 60/391,786  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/392,018  
; PRIOR FILING DATE: 2002-06-26  
; NUMBER OF SEQ ID NOS: 2459  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1034  
; LENGTH: 324  
; TYPE: PRT  
; ORGANISM: Neisseria meningitidis  
US-10-389-566-1034

Query Match 80.0%; Score 36; DB 4; Length 324;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NAENFV 9  
Db 175 NAENFV 181

RESULT 9  
US-10-389-566-2328

; Sequence 2328, Application US/10389566  
; Publication No. US20040025202A1  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC  
; APPLICANT: Laurie, Cathy C  
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants  
; FILE REFERENCE: 38-77(52900)D  
; CURRENT APPLICATION NUMBER: US/10/389,566  
; CURRENT FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US 60/365,301  
; PRIOR FILING DATE: 2002-03-15  
; PRIOR APPLICATION NUMBER: US 60/391,786  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/392,018  
; PRIOR FILING DATE: 2002-06-26  
; NUMBER OF SEQ ID NOS: 2459  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2328  
; LENGTH: 324  
; TYPE: PRT  
; ORGANISM: Neisseria meningitidis  
US-10-389-566-2328

Query Match 80.0%; Score 36; DB 4; Length 324;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NAENFV 9  
Db 175 NAENFV 181

RESULT 10  
US-10-732-923-17878  
; Sequence 17878, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 17878  
; LENGTH: 324  
; TYPE: PRT  
; ORGANISM: Neisseria meningitidis MC58  
US-10-732-923-17878

Query Match 80.0%; Score 36; DB 5; Length 324;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NAENFV 9  
Db 175 NAENFV 181

RESULT 11  
US-10-732-923-17879  
; Sequence 17879, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149

```
; SEQ ID NO 17879
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Neisseria meningitidis Z2491
US-10-732-923-17879

Query Match
Best Local Similarity 80.0%; Score 36; DB 5; Length 324;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEVF 9
Db 175 NAENEVF 181

RESULT 12
US-11-188-298-3960
; Sequence 3960, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 3960
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Neisseria meningitidis MC58
US-11-188-298-3960

Query Match
Best Local Similarity 80.0%; Score 36; DB 6; Length 324;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEVF 9
Db 175 NAENEVF 181

RESULT 13
US-11-188-298-15864
; Sequence 15864, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 15864
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Neisseria meningitidis Z2491
US-11-188-298-15864

Query Match
Best Local Similarity 80.0%; Score 36; DB 6; Length 324;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEVF 9
Db 175 NAENEVF 181

RESULT 14
US-10-467-657-5266
; Sequence 5266, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwin99, version 1.04
; SEQ ID NO 5266
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5266

Query Match
Best Local Similarity 80.0%; Score 36; DB 5; Length 395;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NAENEVF 9
Db 246 NAENEVF 252

RESULT 15
US-10-282-122A-54383
; Sequence 54383, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File wrapper or PALM.
```



```
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54383
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
US-10-282-122A-54383
```

```
Query Match . 80.0%; Score 36; DB 4; Length 583;
Best Local Similarity 77.8%; Pred. No. 2.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 ATNAENEV 9
|:|:|:|:|
Db 61 ASNARNEV 69
```

Search completed: July 5, 2006, 20:51:59  
Job time : 66.8935 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 5.15534 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-8

Perfect score: 45

Sequence: 1 ATNAENEFF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB pep.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	75.6	153	6	US-10-449-902-35066
2	34	75.6	943	6	US-10-449-902-55109
3	33	73.3	469	6	US-10-505-328-457
4	32	71.1	327	6	US-10-471-571A-4214
5	32	71.1	328	6	US-10-449-902-42678
6	31	68.9	98	6	US-10-449-902-30269
7	31	68.9	221	6	US-10-449-902-30681
8	31	68.9	416	6	US-10-471-571A-4782
9	31	68.9	1019	6	US-10-449-902-53317
10	31	68.9	1186	7	US-11-121-154-172
11	30	66.7	170	6	US-10-449-902-46868
12	30	66.7	401	7	US-11-168-010-6
13	30	66.7	401	7	US-11-168-130-5
14	30	66.7	414	6	US-10-449-902-38632
15	30	66.7	424	7	US-11-168-010-5
16	30	66.7	424	7	US-11-168-130-5
17	30	66.7	460	6	US-10-449-902-30876
18	30	66.7	471	6	US-10-511-937-2582
19	30	66.7	471	7	US-11-105-233-67
20	30	66.7	483	7	US-11-105-233-180
21	30	66.7	569	6	US-10-449-902-50324
22	29.5	65.6	357	6	US-10-953-349-13122
23	29.5	65.6	375	6	US-10-953-349-13121
24	29	64.4	23	7	US-11-134-871-2119
25	29	64.4	153	6	US-10-449-902-47002

26	29	64.4	220	6	US-10-953-349-20575	Sequence 20575, A
27	29	64.4	303	6	US-10-953-349-1993	Sequence 1993, Ap
28	29	64.4	322	6	US-10-471-571A-3888	Sequence 3888, Ap
29	29	64.4	344	6	US-10-449-902-37534	Sequence 37534, A
30	29	64.4	377	6	US-10-953-349-1992	Sequence 1992, Ap
31	29	64.4	382	6	US-10-953-349-1991	Sequence 1991, Ap
32	29	64.4	423	6	US-10-449-902-43160	Sequence 43160, A
33	28	62.2	95	6	US-10-953-349-35135	Sequence 35135, A
34	28	62.2	214	6	US-10-953-349-2848	Sequence 2848, Ap
35	28	62.2	217	6	US-10-449-902-44464	Sequence 44464, A
36	28	62.2	220	6	US-10-449-902-30877	Sequence 30877, A
37	28	62.2	220	6	US-10-449-902-33167	Sequence 33167, A
38	28	62.2	232	6	US-10-449-902-41670	Sequence 41670, A
39	28	62.2	242	7	US-11-251-208-190	Sequence 190, App
40	28	62.2	262	7	US-11-174-307B-3930	Sequence 3930, Ap
41	28	62.2	302	6	US-10-953-349-16065	Sequence 16065, A
42	28	62.2	314	6	US-10-471-571A-4458	Sequence 4458, Ap
43	28	62.2	316	6	US-10-953-349-16064	Sequence 16064, A
44	28	62.2	331	6	US-10-471-571A-1962	Sequence 1962, Ap
45	28	62.2	343	6	US-10-953-349-16063	Sequence 16063, A

#### ALIGNMENTS

RESULT 1  
US-10-449-902-35066  
; Sequence 35066, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 35066  
; LENGTH: 153  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-35066  
  
Query Match 75.6%; Score 34; DB 6; Length 153;  
Best Local Similarity 75.0%; Pred. No. 3.3;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 ATNAENEFF 8  
||| |||:  
Db 48 ATNGENDF 55  
  
RESULT 2  
US-10-449-902-55109  
; Sequence 55109, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269

```
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55109
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-55109

Query Match      75.6%; Score 34; DB 6; Length 943;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      1 ATNANEFV 8
Db      838 ATGENDF 845

RESULT 3
US-10-505-928-457
; Sequence 457, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 457
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-505-928-457

Query Match      73.3%; Score 33; DB 6; Length 469;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY      2 TNANEFV 9
Db      188 TAANEFV 195

RESULT 4
US-10-471-571A-4214
; Sequence 4214, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4214
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(327)
; OTHER INFORMATION: thioredoxine reductase
US-10-471-571A-4214

Query Match      68.9%; Score 31; DB 6; Length 98;
Best Local Similarity 75.0%; Pred. No. 8.4;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY      1 ATNANEFV 8

Query Match      71.1%; Score 32; DB 6; Length 327;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 ATNANEFV 9
Db      121 AVNKENDFV 129

RESULT 5
US-10-449-902-42678
; Sequence 42678, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42678
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-42678

Query Match      71.1%; Score 32; DB 6; Length 328;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ATNANEFV 9
Db      190 ATNAENNAV 198

RESULT 6
US-10-449-902-30269
; Sequence 30269, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30269
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-30269

Query Match      68.9%; Score 31; DB 6; Length 98;
Best Local Similarity 75.0%; Pred. No. 8.4;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY      1 ATNANEFV 8
```

Db 39 ATNVENAF 46  
||| ||| |

## RESULT 7

US-10-449-902-30681  
; Sequence 30681, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 30681  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-30681

Query Match 68.9%; Score 31; DB 6; Length 221;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 8  
||| ||| |

Db 165 ATNVENAF 172

## RESULT 8

US-10-471-571A-4782  
; Sequence 4782, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927W0  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 4782  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1)..(416)  
; OTHER INFORMATION: signal recognition particle  
US-10-471-571A-4782

Query Match 68.9%; Score 31; DB 6; Length 416;  
Best Local Similarity 77.8%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 9  
||| ||| |

Db 12 ATNKENEV 20

## RESULT 9

US-10-449-902-53317

; Sequence 53317, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:

; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 53317  
; LENGTH: 1019  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-53317

Query Match 68.9%; Score 31; DB 6; Length 1019;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 9  
||| ||| |

Db 297 AVNAENPFL 305

## RESULT 10

US-11-121-154-172  
; Sequence 172, Application US/111211154  
; Publication No. US20060105914A1  
; GENERAL INFORMATION:  
; APPLICANT: TAYLOR, LARRY EDMUND  
; APPLICANT: WEINER, RONALD M.  
; APPLICANT: HUTCHESON, STEVEN WAYNE  
; APPLICANT: EKBORG, NATHAN A.  
; APPLICANT: HOWARD, MICHAEL  
; TITLE OF INVENTION: PLANT WALL DEGRADATIVE COMPOUNDS AND SYSTEMS  
; FILE REFERENCE: 108172-00121  
; CURRENT APPLICATION NUMBER: US/11/121,154  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,971  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 214  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 172  
; LENGTH: 1186  
; TYPE: PRT  
; ORGANISM: Microbulbifer degradans  
US-11-121-154-172

Query Match 68.9%; Score 31; DB 7; Length 1186;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TNAENEV 8  
||| ||| |

Db 406 TNAANEV 412

## RESULT 11

US-10-449-902-46868  
; Sequence 46868, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.

; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF

; FILE REFERENCE: MOA-A0205Y1-US

; CURRENT APPLICATION NUMBER: US/10/449,902

; CURRENT FILING DATE: 2003-05-29

; PRIOR APPLICATION NUMBER: JP 2002-203269

; PRIOR FILING DATE: 2002-05-30

; PRIOR APPLICATION NUMBER: JP 2002-383870

; PRIOR FILING DATE: 2002-12-11

; NUMBER OF SEQ ID NOS: 56791

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 46868

; LENGTH: 170

; TYPE: PRT

; ORGANISM: Oryza sativa

; US-10-449-902-46868

Query Match 66.7%; Score 30; DB 6; Length 170;

Best Local Similarity 66.7%; Pred. No. 26;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 9

||: ||||

DB 48 ATSGEGEV 56

RESULT 12

US-11-168-010-6

; Sequence 6, Application US/11168010

; Publication No. US20060104962A1

; GENERAL INFORMATION:

; APPLICANT: The Scripps Research Institute

; APPLICANT: Friedlander, Martin

; APPLICANT: Otani, Atsushi

; APPLICANT: Dasilva, Karen

; APPLICANT: Moreno, Stacey (Hanekamp)

; TITLE OF INVENTION: TRANSFECTED HEMATOPOIETIC STEM CELLS AND

; FILE REFERENCE: TSRI-988.3

; CURRENT APPLICATION NUMBER: US/11/168,010

; CURRENT FILING DATE: 2005-06-28

; PRIOR APPLICATION NUMBER: 10/833,743

; PRIOR FILING DATE: 2004-04-28

; PRIOR APPLICATION NUMBER: 10/628,783

; PRIOR FILING DATE: 2003-07-25

; PRIOR APPLICATION NUMBER: 60/467,051

; PRIOR FILING DATE: 2003-05-02

; PRIOR APPLICATION NUMBER: 60/398,522

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 401

; TYPE: PRT

; ORGANISM: homo sapiens

; US-11-168-010-6

Query Match 66.7%; Score 30; DB 7; Length 401;

Best Local Similarity 66.7%; Pred. No. 70;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 9

||: ||||

DB 7 ATEAEEDFV 15

RESULT 13

US-11-168-130-6

; Sequence 6, Application US/11168130

; Publication No. US20060104963A1

; GENERAL INFORMATION:

; APPLICANT: The Scripps Research Institute

; APPLICANT: Friedlander, Martin

; APPLICANT: Otani, Atsushi  
; APPLICANT: Dasilva, Karen  
; APPLICANT: Moreno, Stacey (Hanekamp)  
; TITLE OF INVENTION: TREATMENT OF CONE CELL DEGENERATION WITH  
; FILE REFERENCE: TSRI-988.4  
; CURRENT APPLICATION NUMBER: US/11/168,130  
; CURRENT FILING DATE: 2005-06-28  
; PRIOR APPLICATION NUMBER: 10/933,634  
; PRIOR FILING DATE: 2004-09-03  
; PRIOR APPLICATION NUMBER: 10/833,743  
; PRIOR FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: 10/628,783  
; PRIOR FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: 60/467,051  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: 60/398,522  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: homo sapiens  
; US-11-168-130-6

Query Match 66.7%; Score 30; DB 7; Length 401;

Best Local Similarity 66.7%; Pred. No. 70;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 9

||: ||||

DB 7 ATEAEEDFV 15

RESULT 14

US-10-449-902-38632

; Sequence 38632, Application US/10449902

; Publication No. US20060123505A1

; GENERAL INFORMATION:

; APPLICANT: National Institute of Agrobiological Sciences.

; APPLICANT: Bio-oriented Technology Research Advancement Institution.

; APPLICANT: The Institute of Physical and Chemical Research.

; APPLICANT: Foundation for Advancement of International Science.

; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF

; FILE REFERENCE: MOA-A0205Y1-US

; CURRENT APPLICATION NUMBER: US/10/449,902

; CURRENT FILING DATE: 2003-05-29

; PRIOR APPLICATION NUMBER: JP 2002-203269

; PRIOR FILING DATE: 2002-05-30

; PRIOR APPLICATION NUMBER: JP 2002-383870

; PRIOR FILING DATE: 2002-12-11

; NUMBER OF SEQ ID NOS: 56791

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 38632

; LENGTH: 414

; TYPE: PRT

; ORGANISM: Oryza sativa

; US-10-449-902-38632

Query Match 66.7%; Score 30; DB 6; Length 414;

Best Local Similarity 66.7%; Pred. No. 72;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATNAENEV 9

||: ||||

DB 191 ALRAENDFV 199

RESULT 15

US-11-168-010-5

; Sequence 5, Application US/11168010

; Publication No. US20060104962A1

; GENERAL INFORMATION:  
; APPLICANT: The Scripps Research Institute  
; APPLICANT: Friedlander, Martin  
; APPLICANT: Otani, Atsushi  
; APPLICANT: DaSilva, Karen  
; APPLICANT: Moreno, Stacey (Hanekamp)  
; TITLE OF INVENTION: TRANSFECTED HEMATOPOIETIC STEM CELLS AND  
; TITLE OF INVENTION: METHODS OF TREATMENT OF NEOVASCULAR EYE DISEASES THEREWITH  
; FILE REFERENCE: TSRI-988.3  
; CURRENT APPLICATION NUMBER: US/11/168,010  
; CURRENT FILING DATE: 2005-06-28  
; PRIOR APPLICATION NUMBER: 10/833,743  
; PRIOR FILING DATE: 2004-04-28  
; PRIOR APPLICATION NUMBER: 10/628,783  
; PRIOR FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: 60/467,051  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: 60/398,522  
; PRIOR FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 424  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-11-168-010-5

Query Match 66.7%; Score 30; DB 7; Length 424;  
Best Local Similarity 66.7%; Pred. No. 74;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATNAENEV 9  
||| :||  
Db 30 ATAEEDFV 38

Search completed: July 5, 2006, 20:53:02  
Job time : 6.15534 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 20.2718 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-8

Perfect score: 45

Sequence: 1 ATNAENEV 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/iaa/5 COMB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/iaa/6 COMB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/iaa/7 COMB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/iaa/H COMB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/iaa/PTUS COMB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/iaa/RE COMB.pep.\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/iaa/baCkfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	9	2	US-09-809-003A-8
2	41	91.1	637	2	US-09-949-016-8152
3	41	91.1	643	2	US-09-538-092-844
4	35	77.8	567	2	US-10-104-047-3772
5	34	75.6	271	2	US-09-660-587-46
6	34	75.6	271	2	US-09-811-007A-46
7	33	73.3	251	2	US-09-248-796A-17802
8	33	73.3	398	2	US-10-104-047-3388
9	33	73.3	459	2	US-09-489-039A-12641
10	33	73.3	469	2	US-09-077-606-3
11	33	73.3	476	2	US-09-949-016-9096
12	33	73.3	505	2	US-10-116-821-14
13	33	73.3	546	1	US-09-067-351-1
14	33	73.3	546	2	US-09-360-490-1
15	33	73.3	551	1	US-09-067-351-2
16	33	73.3	551	2	US-09-360-490-2
17	33	73.3	564	2	US-09-949-016-6628
18	33	73.3	569	2	US-09-949-016-11035
19	33	73.3	569	2	US-09-949-016-11036
20	32	71.1	327	2	US-08-956-171E-5208
21	32	71.1	327	2	US-08-781-986A-5208
22	32	71.1	432	2	US-09-270-767-43899
23	32	71.1	459	2	US-09-328-352-4523
24	32	71.1	1091	2	US-09-328-352-5758
25	31	68.9	80	2	US-09-248-796A-15929
26	31	68.9	139	2	US-09-248-796A-20474

```

27 31 68.9 171 2 US-09-134-001C-3566 Sequence 3566, Ap
28 31 68.9 213 1 US-08-531-525-11 Sequence 11, Appl
29 31 68.9 213 1 US-08-718-270A-11 Sequence 11, Appl
30 31 68.9 229 2 US-09-248-796A-24307 Sequence 24307, A
31 31 68.9 305 2 US-09-248-796A-18731 Sequence 18731, A
32 31 68.9 416 2 US-09-007-476-2 Sequence 2, Appli
33 31 68.9 492 1 US-08-463-090B-10 Sequence 10, Appl
34 31 68.9 506 2 US-09-248-796A-15570 Sequence 15570, A
35 31 68.9 574 2 US-10-104-047-2310 Sequence 2310, Ap
36 31 68.9 693 2 US-09-538-092-534 Sequence 534, App
37 31 68.9 1031 2 US-09-914-259-24 Sequence 24, Appl
38 30 66.7 17 2 US-09-979-672-8 Sequence 8, Appli
39 30 66.7 46 2 US-09-813-718-51 Sequence 51, Appl
40 30 66.7 85 2 US-09-813-718-45 Sequence 45, Appl
41 30 66.7 159 2 US-09-270-767-36313 Sequence 36313, A
42 30 66.7 159 2 US-09-270-767-51530 Sequence 51530, A
43 30 66.7 163 2 US-09-248-796A-27537 Sequence 27537, A
44 30 66.7 254 2 US-09-586-106D-101 Sequence 101, App
45 30 66.7 254 2 US-10-799-870-101 Sequence 101, App

```

#### ALIGNMENTS

```

RESULT 1
US-09-809-003A-8
; Sequence 8, Application US/09809003A
; Patent No. 6673351
; GENERAL INFORMATION:
; APPLICANT: O'Daly, Jose
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission
; TITLE OF INVENTION: Psoriasis
; FILE REFERENCE: 302178
; CURRENT APPLICATION NUMBER: US/09/809,003A
; CURRENT FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Leishmania
; US-09-809-003A-8

```

```

Query Match 100.0%; Score 45; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 5e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ATNAENEV 9
Db 1 ATNAENEV 9

```

#### RESULT 2

```

US-09-949-016-8152
; Sequence 8152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8152
; LENGTH: 637

```

```
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8152

Query Match          91.1%; Score 41; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 TNAENEVF 9
Db   278 TNAENEVF 285

RESULT 3
US-09-538-092-844
; Sequence 844, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormatter Version 0.9
; SEQ ID NO 844
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P04264
US-09-538-092-844

Query Match          91.1%; Score 41; DB 2; Length 643;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 TNAENEVF 9
Db   277 TNAENEVF 284

RESULT 4
US-10-104-047-3772
; Sequence 3772, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. 6943241el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3772
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3772

Query Match          77.8%; Score 35; DB 2; Length 567;
Best Local Similarity 87.5%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 TNAENEVF 9
Db   243 TTAENEVF 250

RESULT 5
US-09-660-587-46
; Sequence 46, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 46
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-9 protein
US-09-660-587-46

Query Match          75.6%; Score 34; DB 2; Length 271;
Best Local Similarity 55.6%; Pred. No. 38;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY  1 ATNAENEVF 9
Db   141 ATNSDNKFI 149

RESULT 6
US-09-811-007A-46
; Sequence 46, Application US/09811007A
; Patent No. 6660269
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/811,007A
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/660,587
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 46
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-9 protein
US-09-811-007A-46

Query Match          75.6%; Score 34; DB 2; Length 271;
Best Local Similarity 55.6%; Pred. No. 38;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY  1 ATNAENEVF 9
Db   141 ATNSDNKFI 149

RESULT 7
US-09-248-796A-17802
```

```
; Sequence 17802, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17802
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17802

Query Match      73.3%; Score 33; DB 2; Length 251;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TNAENEVF 9
Db      198 TNPQNEVF 205

RESULT 8
US-10-104-047-3388
; Sequence 3388, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3388
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3388

Query Match      73.3%; Score 33; DB 2; Length 398;
Best Local Similarity 87.5%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TNAENEVF 9
Db      110 TAAENEVF 117

RESULT 9
US-09-489-039A-12641
; Sequence 12641, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12641
```

```
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12641

Query Match      73.3%; Score 33; DB 2; Length 459;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 ATNAENEVF 9
Db      132 STNADNPFV 140

RESULT 10
US-09-077-606-3
; Sequence 3, Application US/09077606
; Patent No. 6774220
; GENERAL INFORMATION:
; APPLICANT: JIANG, Pan Hong
; APPLICANT: KABA, Aboubacar
; APPLICANT: CHANY-FOURNIER, Francoise
; APPLICANT: CERUTTI, Italina
; APPLICANT: CHANY, Charles
; TITLE OF INVENTION: COMPOUNDS HAVING LECTINIC PROPERTIES AND THEIR
; TITLE OF INVENTION: BIOLOGICAL APPLICATIONS
; FILE REFERENCE: 040388/0113
; CURRENT APPLICATION NUMBER: US/09/077,606
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: WO PCT/FR96/01937
; EARLIER FILING DATE: 1996-12-04
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-606-3

Query Match      73.3%; Score 33; DB 2; Length 469;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TNAENEVF 9
Db      188 TAAENEVF 195

RESULT 11
US-09-949-016-9096
; Sequence 9096, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 9096
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9096
```

Query Match 73.3%; Score 33; DB 2; Length 476;  
Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TNAENEV 9  
| | | | |  
Db 195 TAAENEV 202

RESULT 12  
US-10-116-821-14  
; Sequence 14, Application US/10116821  
; Patent No. 7033804  
; GENERAL INFORMATION:  
; APPLICANT: Dodge, Timothy C.  
; APPLICANT: Valle, Fernando  
; TITLE OF INVENTION: Methods for the Production of Products  
; FILE REFERENCE: in Host Cells  
; FILE REFERENCE: GC620-3  
; CURRENT APPLICATION NUMBER: US/10/116,821  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: US 60/282,277  
; PRIOR FILING DATE: 2001-04-05  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 505  
; TYPE: PRT  
; ORGANISM: Pantoea citrea  
US-10-116-821-14

Query Match 73.3%; Score 33; DB 3; Length 505;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 TNAENEV 9  
| | | | |  
Db 2 TNAENKYI 9

RESULT 13  
US-09-067-351-1  
; Sequence 1, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,351  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555

; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PANTUT02  
; CLONE: 1467090  
US-09-067-351-1

Query Match 73.3%; Score 33; DB 1; Length 546;  
Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TNAENEV 9  
| | | | |  
Db 265 TAAENEV 272

RESULT 14  
US-09-360-490-1  
; Sequence 1, Application US/09360490  
; Patent No. 6221843  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/360,490  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/067,351  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PANTUT02  
; CLONE: 1467090  
US-09-360-490-1

Query Match 73.3%; Score 33; DB 2; Length 546;  
Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TNAENEFV 9  
| | | | |  
Db 265 TAAENEFV 272

RESULT 15  
US-09-067-351-2  
; Sequence 2, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,351  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 551 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: KERANOT02  
; CLONE: 2029060  
; US-09-067-351-2

Query Match 73.3%; Score 33; DB 1; Length 551;  
Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TNAENEFV 9  
| | | | |  
Db 247 TAAENEFV 254

Search completed: July 5, 2006, 19:24:25  
Job time : 21.2718 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 19.4175 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928A-9

Perfect score: 66

Sequence: 1 XXYSELNRVQLRLSI 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80:\*

1: Pirl:\*

2: Pirl2:\*

3: Pirl3:\*

4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	83.3	643	1 KRHU2	keratin 1, type II
2	52	78.8	581	1 KRMS2	keratin, type II c
3	49	74.2	524	2 A23518	keratin, 57K type
4	47	71.2	384	2 I61769	keratin 6d, type I
5	47	71.2	553	2 I59009	epidermal keratin
6	47	71.2	564	1 KRHUEA	keratin 6a, type I
7	47	71.2	564	1 KRHUEB	keratin 6b, type I
8	47	71.2	564	2 I61770	keratin 6e, type I
9	47	71.2	564	2 I61768	keratin 6c, type I
10	47	71.2	564	2 I61771	keratin 6f, type I
11	47	71.2	645	2 A44861	keratin, 67K type
12	45	69.7	534	2 I37942	keratin 4, type II
13	45	68.2	629	2 A29666	keratin, 65K type
14	45	68.2	638	2 I53169	cytokeatin 2 - hu
15	44	66.7	590	2 A29904	keratin 5, type II
16	44	66.7	629	2 S42629	keratin K3 - rabbi
17	43	65.2	491	2 S05408	keratin, type II,
18	43	65.2	503	2 S29094	keratin, type II,
19	43	65.2	508	1 KRSHU2	keratin type II, m
20	42	63.6	390	2 A60093	cytokeatin, type
21	42	63.6	469	2 S05602	keratin K7, type I
22	42	63.6	489	2 B24177	keratin, 55K type
23	42	63.6	506	2 S58150	hypotheical-prote
24	42	63.6	520	2 J50291	intermediate filam
25	42	63.6	623	2 I49969	keratin, 58K type
26	41	62.1	513	2 S08381	probable membrane
27	41	62.1	532	2 S54571	keratin-like prote
28	40	60.6	257	2 I38025	conserved hypothe
29	40	60.6	352	2 C83451	

#### ALIGNMENTS

##### RESULT 1

KRHU2

keratin 1, type II, cytoskeletal - human

N;Alternate names: 67K type II epidermal keratin; cytokeratin 1

C;Species: Homo sapiens (man)

C;Date: 04-Dec-1986 #sequence, revision 22-Oct-1999 #text\_change 10-Dec-1999

C;Accession: A22940; A02950; A43342

R;Johnson, L.D.; Idler, W.W.; Zhou, X.M.; Roop, D.R.; Steinert, P.M.

Proc. Natl. Acad. Sci. U.S.A. 82, 1896-1900, 1985

A;Reference number: A22940; MUID:85166239; PMID:2580302

A;Accession: A22940

A;Molecule type: DNA

A;Residues: 1-643 <JOH>

A;Cross-references: UNIPARC:UPI0000173D59; GB:M98776; GB:M11215; GB:M11845; GB:M11846; N

A;Note: translation of initiator Met is not shown

R;Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.

J. Biol. Chem. 260, 7142-7149, 1985

A;Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000

late filament subunits.

A;Reference number: A92535; MUID:85207740; PMID:2581964

A;Accession: A02950

A;Molecule type: mRNA

A;Residues: 151-183, 'K', 185-199, 'M', 201-204, 'K', 206-236, 'S', 238-239, 'R', 241-356, 'Y', 358-

'S', 638-643 <STE>

A;Cross-references: UNIPARC:UPI000016ABD0; GB:M10938; NID:gl86787; PIDN:AAA36153.1; PID:

A;Experimental source: tissue neonatal foreskin

A;Note: the authors translated the codon CUG for residue 476 as Met

R;Chipev, C.C.; Korge, B.P.; Markova, N.; Bale, S.J.; DiGiovanna, J.J.; Compton, J.G.; S

Cell 70, 821-828, 1992

A;Title: A leucine---proline mutation in the H1 subdomain of keratin 1 causes epidermol

A;Reference number: A43342; MUID:92386601; PMID:1381288

A;Accession: A43342

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 144-146, 'P', 148-159, 'P', 161-183, 'K', 185-186 <CHI>

A;Cross-references: UNIPARC:UPI0000173D5A; GB:M98776; GB:M11215; GB:M11845; GB:M11846; N

A;Note: sequence extracted from NCBI backbone (NCBIP:112784)

C;Comment: The cytoskeletal and microfibrillar keratins are classified into two types, t

atin IF protein subunit appears to be a heterotetramer of two type I and two type II pro

C;Comment: Keratin 1 is expressed in terminally differentiating epidermis.

C;Genetics:

A;Gene: GDB:KRT1

A;Cross-references: GDB:128198; OMIM:139350

A;Map position: 12q11-12q13

A;Note: defects in this gene may result in epidermolytic hyperkeratosis

C;Complex: heterotetramer of two type I, usually keratin 10 (see PIR:KRHU0), and two typ

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil; heterotetramer; intermediate filament

F;4-179/Domain: head <HED>

F;4-143/Region: E1 and V1 subdomains

F;14-179/Region: H1 subdomain

F;180-492/Domain: rod <ROD>

hypotheical prote  
keratin, 64K type  
keratin type II, h  
keratin, type II,  
keratin type II -  
hypotheical prote  
mucs (AF238304) [i  
hypotheical prote  
DNA polymerase III  
hypotheical prote  
hypotheical prote  
gonadal protein z6  
probable uracil ph  
glucose-6-phosphat  
guanosine-3',5'-bi  
guanosine-3',5'-bi

30 40 60.6 373 2 A71690  
31 40 60.6 425 1 KRXL2A  
32 40 60.6 479 2 A61368  
33 40 60.6 480 2 A56694  
34 40 60.6 495 2 I57463  
35 40 60.6 532 2 H85035  
36 39 59.1 170 2 F97501  
37 39 59.1 170 2 AB2720  
38 39 59.1 1451 2 D64203  
39 38 57.6 47 2 D95156  
40 38 57.6 77 2 F95154  
41 38 57.6 90 2 A30172  
42 38 57.6 360 2 H71324  
43 38 57.6 558 2 I48073  
44 38 57.6 718 2 B81058  
45 38 57.6 725 2 E81819

F:180-214/Region: coil 1A  
 F:215-226/Region: linker 1  
 F:227-327/Region: coil 1B  
 F:328-344/Region: linker 12  
 F:345-363/Region: linker 2A  
 F:364-371/Region: linker 2  
 F:372-492/Region: coil 2B  
 F:430/Region: stutter  
 F:493-643/Domain: tail <END>  
 F:493-512/Region: H2 subdomain  
 F:513-643/Region: V2 and E2 subdomains

Query Match 83.3%; Score 55; DB 1; Length 643;  
 Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;  
 Matches 12; Conservative 0; Mismatches 0;

QY 4 SELNRVIOQLRS 15  
 |||||  
 Db 398 SELNRVIOQLRS 409  
 |||||

RESULT 2  
 KRMS2  
 keratin, type II cytoskeletal - mouse (fragment)  
 N:Alternate names: 67-kDa type II keratin  
 C:Species: Mus musculus (house mouse)  
 C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004  
 C:Accession: A02951  
 R:Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.  
 J. Biol. Chem. 260, 7142-7149, 1985  
 A:Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000  
 kDa filament subunits  
 A:Reference number: A92535; MUID:85207740; PMID:2581964  
 A:Accession: A02951  
 A:Molecule type: mRNA  
 A:Residues: 1-581 <STE>  
 A:Cross-references: UNIPROT:P04104; UNIPARC:UPI0000173DSB; GB:M10937; NID:g198622; PID:g198622  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; intermediate filament  
 F:1-106/Domain: head (fragment) <HED>  
 F:1-70/Region: V1 subdomain (fragment)  
 F:71-106/Region: H1 subdomain  
 F:107-419/Domain: rod <ROD>  
 F:107-141/Region: coil 1A  
 F:142-153/Region: linker 1  
 F:154-254/Region: coil 1B  
 F:255-271/Region: linker 12  
 F:272-290/Region: coil 2A  
 F:291-298/Region: linker 2  
 F:299-419/Region: coil 2B  
 F:357/Region: stutter  
 F:420-581/Domain: tail <END>  
 F:420-439/Region: H2 subdomain  
 F:440-581/Region: V2 and E2 subdomains

Query Match 78.8%; Score 52; DB 1; Length 581;  
 Best Local Similarity 91.7%; Pred. No. 0.41; Indels 0; Gaps 0;  
 Matches 11; Conservative 1; Mismatches 1;

QY 4 SELNRVIOQLRS 15  
 |||||  
 Db 325 SELNRVIOQLRS 336  
 |||||

RESULT 3  
 A23518  
 keratin, 57K type II cytoskeletal - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 09-Jul-2004  
 C:Accession: A23518  
 R:Knapp, B.; Rentrop, M.; Schweizer, J.; Winter, H.  
 Nucleic Acids Res. 14, 751-763, 1986  
 A:Title: Nonepidermal members of the keratin multigene family: cDNA sequences and in situ

A:Reference number: A93640; MUID:86120369; PMID:2418416  
 A:Accession: A23518  
 A:Molecule type: mRNA  
 A:Residues: 1-524 <KNA>  
 A:Cross-references: UNIPROT:P07744; UNIPARC:UPI00000298E2; GB:X03491; NID:g52784; PIDN:O7744  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil

Query Match 74.2%; Score 49; DB 2; Length 524;  
 Best Local Similarity 83.3%; Pred. No. 1.2; Indels 0; Gaps 0;  
 Matches 10; Conservative 2; Mismatches 0;

QY 4 SELNRVIOQLRS 15  
 |||||  
 Db 365 SELNRVIOQLRA 376  
 |||||

RESULT 4  
 I61769  
 keratin 6d, type II - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
 C:Accession: I61769  
 R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
 J. Biol. Chem. 270, 18581-18592, 1995  
 A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly homologous keratins  
 A:Reference number: A57398; MUID:95355491; PMID:7543104  
 A:Accession: I61769  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-384 <RES>  
 A:Cross-references: UNIPROT:P48667; UNIPARC:UPI000012DBL7; GB:L42602; NID:g914823; GB:L42602; GB:L42608; NID:g914829; GB:L42609; NID:g914830; GB:L42610; NID:g1488252; PIDN:AAB6069  
 C:Genetics: KRT6D  
 A:Gene: KRT6D  
 A:Introns: 72/2; 92/3; 124/3; 179/3; 221/3; 295/2; 307/1  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil

Query Match 71.2%; Score 47; DB 2; Length 384;  
 Best Local Similarity 75.0%; Pred. No. 1.9; Indels 0; Gaps 0;  
 Matches 9; Conservative 3; Mismatches 0;

QY 4 SELNRVIOQLRS 15  
 :|||  
 Db 202 ABINRMIOQLRS 213  
 :|||

RESULT 5  
 I59009  
 epidermal keratin subunit II - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
 R:Steinert, P.M.; Parry, D.A.D.; Raccosin, E.L.; Idler, W.W.; Steven, A.C.; Trus, B.L.; F.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713, 1984  
 A:Title: The complete cDNA and deduced amino acid sequence of a type II mouse epidermal keratin  
 A:Reference number: I59009; MUID:85014838; PMID:6207530  
 A:Accession: I59009  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-553 <RES>  
 A:Cross-references: UNIPROT:P50446; UNIPARC:UPI000016CEB7; GB:K02108; NID:g198634; PIDN:P50446  
 R:Finch, J.; Andrews, K.; Krieg, P.; Fuerstenberger, G.; Slaga, T.; Ootsuyama, A.; Tanooki, H.  
 Carcinogenesis 12, 1519-1522, 1991  
 A:Title: Identification of a cloned sequence activated during multi-stage carcinogenesis  
 A:Reference number: A61205; MUID:91316763; PMID:1713533  
 A:Accession: A61205  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 528-553 <FIN>  
 A:Cross-references: UNIPARC:UPI00001774C8  
 C:Genetics:



N;Alternate names: 56k type II keratin; keratin, cytoskeletal  
C;Species: Homo sapiens (man)  
C;Date: 04-Dec-1986 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: I61767; A02945  
R;Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A;Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A;Reference number: A57398; MUID:95355491; PMID:7543104  
A;Accession: I61767  
A;Status: translated from GB/EMBL/DBDJB  
A;Molecule type: DNA  
A;Residues: 1-564 <RES>  
A;Cross-references: UNIPROT:P04259; UNIPARC:UPI0000167B38; GB:L42584; NID:9908780; GB:L42585; GB:L42590; NID:9908786; GB:L42591; NID:9908787; GB:L42592; NID:9908788; PIDN:AA41768  
C;Tyner, A.L.; Eichman, M.J.; Fuchs, E.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4683-4687, 1985  
A;Title: The sequence of a type II keratin gene expressed in human skin: conservation of  
A;Reference number: A02945; MUID:85270392; PMID:2410904  
A;Accession: A02945  
A;Molecule type: DNA  
A;Residues: 2-88, 'AG', '91-115, 'PA', '118, 'LC', '122-158, 'IG', '161-254, 'V', '256-564 <TYN>  
A;Cross-references: UNIPARC:UPI0000173D58; GB:M11229; GB:L00205; NID:9186714; PIDN:AAA594  
A;Note: initiator Met not shown  
C;Comment: There are two types of cytoskeletal and microfibrillar keratin, I (acidic) and II (basic) bialginate heteropolymers, composed of complexes formed by the aggregation of at least one  
C;Genetics:  
A;Gene: GDB:KRT6B  
A;Cross-references: GDB:128113; OMIM:148042  
A;Map position: 12pter-12qter  
A;Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; intermediate filament  
F;2-163/Domain: head <HED>  
F;2-127/Region: E1 and V1 subdomains  
F;128-163/Region: H1 subdomain  
F;164-476/Domain: rod <ROD>  
F;164-198/Region: coil 1A  
F;199-210/Region: linker 1  
F;211-311/Region: coil 1B  
F;312-328/Region: linker 12  
F;329-347/Region: coil 2A  
F;348-355/Region: linker 2  
F;356-476/Region: coil 2B  
F;414/Region: stutler  
F;477-564/Domain: tail <END>  
F;477-496/Region: H2 subdomain  
F;497-564/Region: V2 and E2 subdomains

Query Match	71.2%;	Score 47;	DB 1;	Length 564;
Best Local Similarity	75.0%;	Pred. No. 2.9;		
Matches	9;	Conservative 3;	Mismatches 0;	Indels 0;
Gaps	0;			

Qy 4 SELNRVITQLRLS 15  
: : : : :  
Db 382 AEINRMITQLRLS 393

RESULT 8  
I61770  
keratin 6e, type II - human  
C;Species: Homo sapiens (man)  
C;Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
C;Accession: I61770  
R;Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995  
A;Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
A;Reference number: A57398; MUID:95355491; PMID:7543104  
A;Accession: I61770  
A;Status: translated from GB/EMBL/DBDJB  
A;Molecule type: mRNA  
A;Residues: 1-564 <RES>  
A;Cross-references: UNIPROT:P48668; UNIPARC:UPI0000167E31; GB:L42611; NID:9908802; PIDN:AA41768  
C;Genetics:

A:Gene: KRT6E  
 A:Note: this may not be a distinct gene  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil

Query Match 71.2%; Score 47; DB 2; Length 564;  
 Best Local Similarity 75.0%; Pred. No. 2.9;  
 Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRS 15  
 Db 382 AEINRMIOQLRS 393  
 :|:|:|:|:|:|

## RESULT 9

keratin 6c, type II - human  
 C:Species: Homo sapiens (man)  
 C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
 C:Accession: I61768  
 R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
 J. Biol. Chem. 270, 18581-18592, 1995  
 A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly  
 A:Reference number: A57398; MUID:95355491; PMID:7543104  
 A:Accession: I61768  
 A:Status: translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-564 <RES>  
 A:Cross-references: UNIPROT:P48666; UNIPARC:UPI0000167E0; GB:L42593; NID:g908791; GB:L4  
 96; GB:L42599; NID:g908797; GB:L42600; NID:g908798; GB:L42601; NID:g908799; PIDN:AAAC4176  
 C:Genetics:  
 A:Gene: KRT6C  
 A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil

Query Match 71.2%; Score 47; DB 2; Length 564;  
 Best Local Similarity 75.0%; Pred. No. 2.9;  
 Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRS 15  
 Db 382 AEINRMIOQLRS 393  
 :|:|:|:|:|:|

Query Match 71.2%; Score 47; DB 2; Length 564;  
 Best Local Similarity 75.0%; Pred. No. 2.9;  
 Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRS 15  
 Db 382 AEINRMIOQLRS 393  
 :|:|:|:|:|:|

Query Match 71.2%; Score 47; DB 2; Length 564;  
 Best Local Similarity 75.0%; Pred. No. 2.9;  
 Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRS 15  
 Db 382 AEINRMIOQLRS 393  
 :|:|:|:|:|:|

## RESULT 11

A4861  
 keratin, 67K type II epidermal - human  
 N:Alternate names: cytokeratin 2, CK 2; epidermal cytokeratin 2  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
 C:Accession: A4861  
 R:Collin, C.; Moll, R.; Kubicka, S.; Ouhayoun, J.P.; Franke, W.W.  
 Exp. Cell Res. 202, 132-141, 1992  
 A:Title: Characterization of human cytokeratin 2, an epidermal cytoskeletal protein synt  
 A:Reference number: A44861; MUID:92380238; PMID:1380918  
 A:Accession: A44861  
 A:Molecule type: mRNA  
 A:Residues: 1-645 <COL>  
 A:Cross-references: UNIPROT:P35908; UNIPARC:UPI000012DB03; GB:S43646  
 A:Experimental source: epidermis  
 A:Note: the authors translated the codon GGC for residue 146 as Cys  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; intermediate filament

## RESULT 12

Query Match 71.2%; Score 47; DB 2; Length 645;  
 Best Local Similarity 90.9%; Pred. No. 3.4;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLR 14  
 Db 403 SELNRVIOQLR 413  
 :|:|:|:|:|:|

## RESULT 13

Query Match 71.2%; Score 47; DB 2; Length 645;  
 Best Local Similarity 90.9%; Pred. No. 3.4;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## RESULT 14

Query Match 71.2%; Score 47; DB 2; Length 645;  
 Best Local Similarity 90.9%; Pred. No. 3.4;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## RESULT 15

Query Match 71.2%; Score 47; DB 2; Length 645;  
 Best Local Similarity 90.9%; Pred. No. 3.4;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## RESULT 16

Query Match 71.2%; Score 47; DB 2; Length 645;  
 Best Local Similarity 90.9%; Pred. No. 3.4;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## RESULT 17

Query Match 71.2%; Score 47; DB 2; Length 645;  
 Best Local Similarity 90.9%; Pred. No. 3.4;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

F:465-534/Domain: tail #status predicted <TAI>

Query Match 69.7%; Score 46; DB 2; Length 534;  
Best Local Similarity 75.0%; Pred. No. 4.1;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRLS 15

Db 370 AELNRMIQLRLA 381

## RESULT 13

A29666

keratin, 65K type II cytoskeletal - human

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 20-Apr-2001

C;Accession: A29666

R;Klinge, E.M.; Sylvestre, Y.R.; Freedberg, I.M.; Blumenberg, M.

J. Mol. Evol. 24, 319-329, 1987

A;Title: Evolution of keratin genes: different protein domains evolve by different paths

A;Reference number: A29666; MUID:87254239; PMID:2439698

A;Accession: A29666

A;Molecule type: DNA

A;Residues: 1-629 <KLI>

A;Cross-references: UNIPARC:UPI0000161C35; GB:X05418; NID:934040; PIDN:CAA28991.1; PID:922; PIDN:CAA28994.1; PID:g1335193; PIDN:CAA28995.1; PID:g1335194; GB:X05421; NID:934047;

A;Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 2

A;Note: the complete translation is not annotated in GenBank entries HSKR65A, HSKR65B,

s mistranslated as a Met initiator codon

C;Genetics:

A;Introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1

A;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil

## Query Match

Best Local Similarity 68.2%; Score 45; DB 2; Length 629;

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ELNRVIOQLRLS 15

Db 420 ELNRMIQLRLA 430

## RESULT 14

I53169

cyto keratin 2 - human

C;Species: Homo sapiens (man)

C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004

C;Accession: I53169

R;Collin, C.; Ouhayoun, J.

Differentiation 51, 137-148, 1992

A;Title: Suprabasal marker proteins distinguishing keratinizing squamous epithelia: Cyt

A;Reference number: I53169; MUID:93114504; PMID:1282112

A;Accession: I53169

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 1-638 <RES>

A;Cross-references: UNIPROT:Q01546; UNIPARC:UPI000012DB04; GB:M99063; NID:g181389; PIDN:

C;Superfamily: cytoskeletal keratin

## Query Match

Best Local Similarity 68.2%; Score 45; DB 2; Length 638;

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ELNRVIOQLRLS 15

Db 403 ELNRMIQLRLA 413

## RESULT 15

A29904

keratin 5, type II, epidermal - human

N;Alternate names: 58K type II keratin; cyto keratin 5; K5

C;Species: Homo sapiens (man)  
C;Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 09-Jul-2004  
C;Accession: A29904; A60865; A32568; A27679; I58080  
R;Eckert, R.L.; Rorke, E.A.  
DNA 7, 337-345, 1988

A;Title: The sequence of the human epidermal 58-kD (S) type II keratin reveals an absenc

A;Reference number: A29904; MUID:88296082; PMID:2456903

A;Accession: A29904

A;Molecule type: DNA; mRNA

A;Residues: 1-590 <ECK>

A;Cross-references: UNIPROT:P13647; UNIPARC:UPI000012DB0D; GB:M21389; NID:g186697; PIDN:

R;Galup, C.; Darmon, M.Y.

J. Invest. Dermatol. 91, 39-42, 1988

A;Title: Isolation and characterization of a cDNA clone coding for human epidermal kerat

A;Reference number: A60865; MUID:88258089; PMID:2455002

A;Accession: A60865

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 348-374, 'E', 376-386, 'S', 388-436, 438-542, 'S', 544-590 <GAL>

A;Cross-references: UNIPARC:UPI00001774BD

R;Lersch, R.; Stellmach, V.; Stocks, C.; Giudice, G.; Fuchs, E.

Mol. Cell. Biol. 9, 3685-3697, 1989

A;Title: Isolation, sequence, and expression of a human keratin K5 gene: transcriptional

A;Reference number: A32568; MUID:89384595; PMID:2476664

A;Accession: A32568

A;Molecule type: DNA

A;Residues: 2-8, 'SGA', 12-36, 'GP', 39-78, 'S', 80-196, 'E', 198-260, 'Q', 262-270, 'H', 272-386, 'S',

A;Cross-references: UNIPARC:UPI00001774BB; GB:M28496; NID:g186728

A;Note: the authors translated the codon CGT for residue 15 as Pro and GAG for residue 2

A;Note: this translation is not annotated in GenBank entry HUMKERK5X, release 111.0

R;Lersch, R.; Fuchs, E.

Mol. Cell. Biol. 8, 486-493, 1988

A;Title: Sequence and expression of a type II keratin, K5, in human epidermal cells.

A;Reference number: A27679; MUID:88094424; PMID:2447486

A;Accession: A27679

A;Molecule type: mRNA

A;Residues: 83-196, 'E', 198-260, 'Q', 262-270, 'H', 272-386, 'S', 388-557, 'S', 559-590 <LE2>

A;Cross-references: UNIPARC:UPI000016ABA6; GB:M19723; NID:g186726; PIDN:AAA36145.1; PID:

R;Lane, E.B.; Rugg, E.L.; Navsaria, H.; Leigh, I.M.; Heagerty, A.H.W.; Ishida-Yamamoto,

Nature 356, 244-246, 1992

A;Title: A mutation in the conserved helix termination peptide of keratin 5 in hereditar

A;Reference number: I58080; MUID:92204232; PMID:1372711

A;Accession: I58080

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 467-474, 'G', 476-481 <LAN>

A;Cross-references: UNIPARC:UPI000011F7CB; GB:S90014; NID:g247707; PIDN:AAB21859.1; PID:

C;Comment: This type II keratin is expressed together with the type I keratin 14 in the

C;Genetics:

A;Gene: GDB:KRT5; EBS2

A;Cross-references: GDB:128110; OMIM:148040

A;Map position: 12q11-12q13

A;Introns: 185/3; 257/2; 277/3; 309/3; 364/3; 406/3; 480/2; 492/1

A;Note: defects in this gene may result in Dowling-Meara type epidermolysis bullosa simp

C;Complex: heterotetramer of two type I, usually keratin 14 (see PIR:KGHUE), and two type

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil; cytoskeleton; epidermis; heterotetramer

Query Match 66.7%; Score 44; DB 2; Length 590;

Best Local Similarity 66.7%; Pred. No. 10;

Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRLS 15

Db 387 TENRMIQLRLA 398

Search completed: July 5, 2006, 19:20:32

Job time : 20.4175 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 163.728 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-9

Perfect score: 66

Sequence: 1 XXYSELNRVQLRLRSI 16

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	83.3	625	Q6IMF3_RAT	Q6imf3 rattus norv
2	55	83.3	643	1 K2C6B_HUMAN	P04264 homo sapien
3	52	78.8	627	1 K2C1_MOUSE	P04104 mus musculu
4	52	78.8	637	Q9D2R8_MOUSE	Q9d2k8 mus musculu
5	50	75.8	79	Q63280_RAT	Q63280 rattus norv
6	50	75.8	381	Q99MH7_MOUSE	Q99mh7 mus musculu
7	50	75.8	398	Q8NA87_HUMAN	Q8na87 homo sapien
8	50	75.8	510	Q8WWZ0_HUMAN	Q8wwz0 homo sapien
9	50	75.8	511	Q8WWY9_HUMAN	Q8wwy9 homo sapien
10	50	75.8	536	1 K2C4_RAT	Q6ig00 rattus norv
11	50	75.8	551	Q8BGZ7_MOUSE	Q8bgz7 m 10 days n
12	50	75.8	552	Q4FZU2_RAT	Q4fzu2 rattus norv
13	50	75.8	573	Q6IG05_RAT	Q6ig05 rattus norv
14	50	75.8	576	1 K2C5_RAT	Q6p6q2 rattus norv
15	50	75.8	580	1 K2C5_MOUSE	Q922u2 mus musculu
16	50	75.8	601	1 K2C5_BOVIN	Q5xqm5 bos taurus
17	50	75.8	606	Q32P04_MOUSE	Q32p04 mus musculu
18	50	75.8	619	Q6EIY9_CANFA	Q6eiy9 canis famil
19	49	74.2	525	1 K2C4_MOUSE	P07744 mus musculu
20	48	72.7	532	Q6IFZ4_RAT	P61fz4 rattus norv
21	48	72.7	538	Q8BI82_MOUSE	Q8bi82 mus musculu
22	47	71.2	203	Q9CV13_MOUSE	Q9cv13 mus musculu
23	47	71.2	278	Q9CV72_MOUSE	Q9cv72 mus musculu
24	47	71.2	384	1 K2C6D_HUMAN	P48667 homo sapien
25	47	71.2	551	Q95678_HUMAN	Q95678 homo sapien
26	47	71.2	551	Q9NSA9_HUMAN	Q9nsa9 homo sapien
27	47	71.2	552	1 K2C6A_MOUSE	P50446 mus musculu
28	47	71.2	552	Q6IFZ8_MOUSE	Q6ifz8 mus musculu
29	47	71.2	554	Q3UV11_MOUSE	Q3uv11 mus musculu
30	47	71.2	561	1 K2C6B_MOUSE	Q92331 mus musculu
31	47	71.2	563	1 K2C6A_HUMAN	P02538 homo sapien

32	47	71.2	563	1 K2C6B_HUMAN	P04259 homo sapien
33	47	71.2	563	1 K2C6C_HUMAN	P48666 homo sapien
34	47	71.2	563	1 K2C6E_HUMAN	P48668 homo sapien
35	47	71.2	564	2 Q2TAZ9_HUMAN	Q2ta29 homo sapien
36	47	71.2	590	1 K2C5_HUMAN	P13647 homo sapien
37	47	71.2	590	2 Q5XGT4_XENLA	Q5xgt4 xenopus lae
38	47	71.2	639	2 Q4VAQ2_HUMAN	Q4vaq2 homo sapien
39	47	71.2	645	1 K22E_HUMAN	P35908 homo sapien
40	46	69.7	381	2 Q32MB2_HUMAN	Q32mb2 homo sapien
41	46	69.7	534	1 K2C4_HUMAN	P19013 homo sapien
42	46	69.7	539	2 Q6NXH9_MOUSE	Q6nxh9 mus musculu
43	46	69.7	540	2 Q86Y46_HUMAN	Q86y46 homo sapien
44	46	69.7	553	2 Q6IG03_RAT	Q6ig03 rattus norv
45	46	69.7	559	2 Q6PVZ5_CHICK	Q6pvz5 gallus gall

#### ALIGNMENTS

```
RESULT 1
Q6IMF3_RAT      PRELIMINARY; PRT; 625 AA.
ID Q6IMF3_RAT
AC Q6IMF3;
DT 05-JUL-2004, integrated into UniprotKB/TREMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Type II Keratin Kbl.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;
RA Hesse M., Zimek A., Weber K., Magin T.M.;
RT "Comprehensive analysis of keratin gene clusters in humans and
   rodents.";
RL Eur. J. Cell Biol. 83:19-26(2004).
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
   EMBL/GenBank/DBJ third party annotation (TPA) entry.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; BK001580; DAA02055.1; -; mRNA.
DR Ensembl; ENSRNOG0000028996; Rattus norvegicus.
DR GO; GO:0005982; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin I.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR18993.SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 625 AA; 64831 MW; 790AAC59E2A2707B CRC64;

Query Match      83.3%; Score 55; DB 2; Length 625;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 SELNRVQLRLRS 15
Db 398 SELNRVQLRLRS 409

RESULT 2
K2C1_HUMAN
ID K2C1_HUMAN      STANDARD; PRT; 643 AA.
```

AC DT 20-MAR-1987, integrated into UniProtKB/Swiss-Prot.  
 DT 01-FEB-1996, sequence version 4.  
 DT 07-FEB-2006, entry version 16.  
 DE Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1)  
 GN (kl1) (67 kDa cytokeratin) (Hair alpha protein).  
 DN Names=KRT1; Synonyms=KRTA;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
 RX MEDLINE=85166239; PubMed=2580302;  
 RA Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;  
 RT "Structure of a gene for the human epidermal 67-kDa keratin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
 RA Whitlock N.V., Eady R.A.J., McGrath J.A.;  
 RT "Genomic organization of the human keratin 1 gene.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
 RA Hatsell S.J., Eady R.A.J., Wennerstrand L., Dopping-Hepenstal P.J.,  
 RA Leigh I.M., Munro C., Kelsell D.P.;  
 RT "Novel splice site mutation in keratin 1 underlies mild epidermolytic  
 RT palmoplantar keratoderma in three kindreds.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA], AND VARIANT TYR-357.  
 RC TISSUE=skin;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP PRELIMINARY NUCLEOTIDE SEQUENCE [MRNA] OF 151-643.  
 RX MEDLINE=85207740; PubMed=2581964;  
 RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,  
 RA Roop D.R.;  
 RT "Amino acid sequences of mouse and human epidermal type II keratins of  
 RT Mr 67,000 provide a systematic basis for the structural and functional  
 RT diversity of the end domains of keratin intermediate filament  
 RT subunits.";  
 RL J. Biol. Chem. 260:7142-7149(1985).  
 RN [6]  
 RP PROTEIN SEQUENCE OF 1-7.  
 RC TISSUE=platelet;  
 RX MEDLINE=22608298; PubMed=12665801; DOI=10.1038/nbt810;  
 RA Gevaert K., Goethals M., Martens L., Van Damme J., Staes A.,  
 RA Thomas G.R., Vandekerckhove J.;  
 RT "Exploring proteomes and analyzing protein processing by mass  
 RT spectrometric identification of sorted N-terminal peptides.";  
 RL Nat. Biotechnol. 21:566-569(2003).  
 RN [7]  
 RP CITRULLINATION  
 RX MEDLINE=96374388; PubMed=8780679; DOI=10.1006/bbrc.1996.1240;  
 RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;  
 RT "Preferential demination of keratin K1 and filaggrin during the  
 RT terminal differentiation of human epidermis.";  
 RL Biochem. Biophys. Res. Commun. 225:712-719(1996).  
 RN [8]  
 RP VARIANT EHK PRO-160.  
 RX MEDLINE=92386601; PubMed=1381288; DOI=10.1016/0092-8674(92)90315-4;  
 RA Chipev C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,  
 RA Compton J.G., Steinert P.M.;  
 RT "A leucine-->proline mutation in the H1 subdomain of keratin 1 causes  
 RT epidermolytic hyperkeratosis.";  
 RL Cell 70:821-828(1992).  
 RN [9]  
 RP VARIANT ALLELE 1B 559-GLY--TYR-565 DEL.  
 RX MEDLINE=93107743; PubMed=1281859; DOI=10.1111/1523-1747.ep12614149;  
 RA Korge B.P., Compton J.G., Steinert P.M., Mischke D.;  
 RT "The two size alleles of human keratin 1 are due to a deletion in the  
 RT glycine-rich carboxyl-terminal V2 subdomain.";  
 RL J. Invest. Dermatol. 99:697-702(1992).  
 RN [10]  
 RP VARIANT EHK GLN-489.  
 RX MEDLINE=92376531; PubMed=1380725;  
 RA Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,  
 RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;  
 RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic  
 RT hyperkeratosis.";  
 RL Science 257:1128-1130(1992).  
 RN [11]  
 RP VARIANT EHK CYS-481.  
 RX MEDLINE=94216497; PubMed=7512983;  
 RA Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;  
 RT "Genetic mutations in the K1 and K10 genes of patients with  
 RT epidermolytic hyperkeratosis. Correlation between location and disease  
 RT severity.";  
 RL J. Clin. Invest. 93:1533-1542(1994).  
 RN [12]  
 RP VARIANTS EHK GLY-154; SER-187 AND PRO-192.  
 RX MEDLINE=94117869; PubMed=7507151; DOI=10.1111/1523-1747.ep12371725;  
 RA Yang J.-M., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,  
 RA Steinert P.M., Compton J.G.;  
 RT "Mutations in the H1 and 1A domains in the keratin 1 gene in  
 RT epidermolytic hyperkeratosis.";  
 RL J. Invest. Dermatol. 102:17-23(1994).  
 RN [13]  
 RP VARIANTS EHK PRO-185 AND SER-187.  
 RX MEDLINE=94117870; PubMed=7507152; DOI=10.1111/1523-1747.ep12371726;  
 RA McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,  
 RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,  
 RA Morley S.M.;  
 RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous  
 RT congenital ichthyosiform erythroderma (BCIE).";  
 RL J. Invest. Dermatol. 102:24-30(1994).  
 RN [14]  
 RP VARIANT NEPPK ILE-73.  
 RX MEDLINE=95096501; PubMed=7528239; DOI=10.1111/1523-1747.ep12412771;  
 RA Kimonis V., Digiovanna J.J., Yang J.-M., Doyle S.Z., Bale S.J.,  
 RA Compton J.G.;  
 RT "A mutation in the V1 end domain of keratin 1 in non-epidermolytic  
 RT palmar-plantar keratoderma.";  
 RL J. Invest. Dermatol. 103:764-769(1994).  
 RN [15]  
 RP VARIANT EHK VAL-339.  
 RX MEDLINE=99072666; PubMed=9856846;  
 RX DOI=10.1046/j.1523-1747.1998.00389.x;  
 RA Kremer H., Lavrijssen A.P., McLean W.H.I., Lane E.B., Melchers D.,  
 RA Ruiter D.J., Mariman E.C., Steijlen P.M.;  
 RT "An atypical form of bullous congenital ichthyosiform erythroderma is  
 RT caused by a mutation in the L12 linker region of keratin 1.";  
 RL J. Invest. Dermatol. 111:1224-1226(1998).  
 RN [16]

```
RP VARIANTS AEI PHE-478 AND THR-478.
RX MEDLINE=99162195; PubMed=10053007;
RA Sybert V.P., Francis J.S., Corden L.D., Smith L.T., Weaver M.,
RA Stephens K., McLean W.H.I.;
RT "Cyclic ichthyosis with epidermolytic hyperkeratosis: a phenotype
RT conferred by mutations in the 2B domain of keratin K1.",
RL Am. J. Hum. Genet. 64:732-738(1999).
RN [17]
RP VARIANT EHK THR-187.
RX MEDLINE=99247363; PubMed=10232403;
RA Arin M.J., Longley M.A., Kuster W., Huber M., Hohl D., Rothnagel J.A.,
RA Roop D.R.;
RT "An asparagine to threonine substitution in the 1A domain of keratin
RT 1: a novel mutation that causes epidermolytic hyperkeratosis.",
RL Exp. Dermatol. 8:124-127(1999).
RN [18]
RP VARIANT AEI PHE-478.
RX MEDLINE=20662276; PubMed=10597140;
RA Michael E.J., Schneiderman P., Grossman M.E., Christiano A.M.;
RT "Epidermolytic hyperkeratosis with polycyclic psoriasiform plaques
RT resulting from a mutation in the keratin 1 gene.",
RL Exp. Dermatol. 8:501-503(1999).
RN [19]
RP VARIANT EHK PRO-213.
RX MEDLINE=20305470; PubMed=10844506;
RA DOI=10.1046/j.1365-2230.2000.00625.x;
RA Cserhalmi-Friedman P.B., Squeo R., Gordon D., Garzon M.,
RA Schneiderman P., Grossman M.E., Christiano A.M.;
RT "Epidermolytic hyperkeratosis in a Hispanic family resulting from a
RT mutation in the keratin 1 gene.",
RL Clin. Exp. Dermatol. 25:241-243(2000).
RN [20]
RP VARIANT EHK THR-478.
RX MEDLINE=20151078; PubMed=10698370;
RA DOI=10.1034/j.1600-0625.2000.009001016.x;
RA Arin M.J., Longley M.A., Epstein E.H. Jr., Rothnagel J.A., Roop D.R.;
RT "Identification of a novel mutation in keratin 1 in a family with
RT epidermolytic hyperkeratosis.",
RL Exp. Dermatol. 9:16-19(2000).
RN [21]
RP VARIANT EHK ASP-154.
RX MEDLINE=21423058; PubMed=11531804;
RA DOI=10.1046/j.1365-2133.2001.04327.x;
RA Whittock N.V., Ashton G.H.S., Griffiths W.A.D., Eady R.A.J.,
RA McGrath J.A.;
RT "New mutations in keratin 1 that cause bullous congenital
RT ichthyosiform erythroderma and keratin 2e that cause ichthyosis
RT bullosa of Siemens.",
RL Br. J. Dermatol. 145:330-335(2001).
RN [22]
RP INVOLVEMENT IN IHCM.
RX MEDLINE=21184521; PubMed=11286616;
RA DOI=10.1046/j.1523-1747.2001.01292.x;

Query Match 83.3%; Score 55; DB 1; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.95; Indels 0; Gaps 0;
Matches 12; Conservative 0; Mismatches 0;

QY 4 SELNRVQLRLS 15
Db 398 SELNRVQLRLS 409
|||||:|||||

RESULT 3
K2C1 MOUSE STANDARD; PRT; 627 AA.
AC P04164;
DT 01-NOV-1986, integrated into UniProtKB/Swiss-Prot.
DT 30-MAY-2000, sequence version 2.
DE 07-FEB-2006, entry version 50.
DE Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1)
DE (K1) (67 kDa cytokeratin).
GN Name=Krt1; Synonyms=Krt2-1;
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=85207740; PubMed=2581964;
RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
RA Roop D.R.;
RT "Amino acid sequences of mouse and human epidermal type II keratins of
RT Mr 67,000 provide a systematic basis for the structural and functional
RT diversity of the end domains of keratin intermediate filament
RT subunits.",
RL J. Biol. Chem. 260:7142-7149(1985).
RN [2]
RP SEQUENCE REVISION.
RX ROOP D.R.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Heterotetramer of two type I and two type II keratins.
CC -1- PTM: Undergoes deamination of some arginine residues
CC (citrullination) (By similarity).
CC -1- MISCELLANEOUS: There are two types of cytoskeletal and
CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
CC 55 and 56-70 kilodaltons, respectively).
CC -1- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: M10937; AAD05191.1; -; mRNA.
DR PIR: A02951; KMS2.
DR HSP; P08670; IGK7.
DR SWISS-2DPAGE; P04104; MOUSE.
DR Ensembl: ENSMUSG00000046834; Mus musculus.
DR MGI: MGI:96698; Krt2-1.
DR InterPro: IPR001664; IF.
DR InterPro: IPR002957; Keratin_I.
DR InterPro: IPR003054; Keratin_II.
DR PANTHER: PTHR1893:SP5; Keratin_II; 1.
DR Pfam: PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Citrullination; Coiled coil; Intermediate filament; Keratin.
FT INIT MET 0 0 By similarity.
FT CHAIN 1 627 Keratin, type II cytoskeletal 1.
FT /FTID=PRO_0000063710.
FT REGION 1 186 Head.
FT REGION 187 496 Rod.
FT REGION 187 222 Coiled coil 1A.
FT REGION 223 242 Linker 1.
FT REGION 243 333 Coiled coil 1B.
FT REGION 334 357 Linker 12.
FT REGION 358 496 Coiled coil 2.
FT REGION 497 627 Tail.
FT SITE 451 451 Stutter.
SQ SEQUENCE 627 AA; 65092 MW; EF7E848654539578 CRC64;

Query Match 78.8%; Score 52; DB 1; Length 627;
Best Local Similarity 91.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLS 15
Db 406 SELNRVQLRLS 417
|||||:|||||

RESULT 4
Q9D2K8 MOUSE PRELIMINARY; PRT; 637 AA.
AC Q9D2K8;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
```

DT 01-JUN-2001, sequence version 1.  
DT 07-FEB-2006, entry version 23.  
DE 0 day neonate head cDNA, RIKEN full-length enriched library,  
DE clone:483436c19 product:keratin complex 2, basic, gene 1, full insert  
DE sequence.  
GN Name=Krt2-1;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Mus.  
ON NCBI\_TaxID=10090;  
RX STRAIN=C57BL/6J; TISSUE=Head;  
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RC PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler K., Aturaliya R.N., Bailey T.L.,  
RA Banael M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA Di Bernharo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Heminger P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Humaneck L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Motagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nakason R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugtara K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kaashina T., Koike M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plesky C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RC PubMed=16141073; DOI=10.1126/science.1112009;  
RX RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RX

OKazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
Nikaido I., Ooato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
Saldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
Schrimal L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
Blake J.A., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
Maggiott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
Hara A., Hashizume W., Imotani K., Sasaki D., Shibata K., Shinagawa A.,  
Miyazaki A., Sakai K., Sasaki D., Ishii Y., Itoh M., Kagawa I.,  
Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Saldarelli R., Bash G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=20499374; PubMed=11012159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-capper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RC MEDLINE=20330913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hazada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
OKazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;



RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]

CC NUCLEOTIDE SEQUENCE.  
 CC STRAIN=C57BL/6J; TISSUE=Head;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami H., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: Belongs to the intermediate filament family.  
 CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----

CC EMBL; AK019521; BAB31776.1; -; mRNA.  
 DR HSP; P08670; ICK7.  
 DR Ensembl; ENSMUSG00000046834; Mus musculus.  
 DR MGI; MGI:96698; Krt2-1.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin I.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR1893:SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Keratin.  
 SQ SEQUENCE 637 AA; 65627 MW; D6344FDD0468EAC7 CRC64;

Query Match 78.8%; Score 52; DB 2; Length 637;  
 Best Local Similarity 91.7%; Pred. No. 3.2;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVQLRLS 15  
 Db 407 SELNRMIQLRLS 418  
 |||||:|||||

RESULT 5  
 Q63280 RAT PRELIMINARY; PRT; 79 AA.  
 ID Q63280 RAT  
 AC Q63280;  
 DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.  
 DT 01-NOV-1996, sequence version 1.  
 DT 07-FEB-2006, entry version 21.  
 DE Keratin K5 (fragment).  
 GN Name=Keratin K5;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidae; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 CC NUCLEOTIDE SEQUENCE.  
 CC STRAIN=Dark agouti; TISSUE=Mammary;  
 RX MEDLINE=93046184; PubMed=1384967;  
 RA Paine M.L., Gibbins J.R., Chew K.E., Demetrious A., Kefferd R.F.;  
 RT "Loss of keratin expression in anaplastic carcinoma cells due to  
 posttranscriptional down-regulation acting in trans."  
 RL Cancer Res. 52:6603-6611(1992).  
 CC -----

Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----

CC EMBL; M89644; AAA99993.1; -; Genomic DNA.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR1893:SF5; Keratin\_II; 1.  
 KW Pfam; PF00038; Filament; 1.  
 KW Keratin.  
 FT NON TER 1 1  
 FT NON TER 79 79  
 SQ SEQUENCE 79 AA; 9395 MW; 92786203021AE926 CRC64;

Query Match 75.8%; Score 50; DB 2; Length 79;  
 Best Local Similarity 83.3%; Pred. No. 0.8;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVQLRLS 15  
 Db 60 SELNRMIQLRLS 71  
 |||||:|||||

RESULT 6  
 Q99MH7 MOUSE PRELIMINARY; PRT; 381 AA.  
 ID Q99MH7 MOUSE  
 AC Q99MH7;  
 DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 20.  
 DE Cytokeratin KRT2-6HF (Fragment).  
 GN Name=Krtcap1; Synonyms=Krt2-6hf;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 CC NUCLEOTIDE SEQUENCE.  
 CC STRAIN=C3H/HeN;  
 RX MEDLINE=2286385; PubMed=1239393;  
 RA Poltner C., Yoshiki A., Fujiwara K., Guenet J.-L., Kusakabe M.;  
 RT "Hague (Hag): a new mouse hair mutation with an unstable semidominant  
 allele.";  
 RL Genetics 162:831-840(2002).  
 CC -1- SIMILARITY: Belongs to the intermediate filament family.  
 CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----

CC EMBL; AF343088; AAK17206.1; -; mRNA.  
 DR HSP; P08670; ICK4.  
 DR Ensembl; ENSMUSG00000022986; Mus musculus.  
 DR MGI; MGI:1923500; Krtcap1.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin I.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR1893:SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Keratin.  
 FT NON TER 1 1  
 SQ SEQUENCE 381 AA; 42383 MW; 4143A882F7F6E9A1 CRC64;

Query Match 75.8%; Score 50; DB 2; Length 381;  
 Best Local Similarity 83.3%; Pred. No. 4.2;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVQLRLS 15

```
Db 199 SEMNRVQLRLS 210
||||:|||||
Best Local Similarity 83.3%; Pred. No. 4, 4;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLS 15
Db 231 SELNRLQIRIS 242
||||:|||||

RESULT 7
Q8NA87 HUMAN PRELIMINARY; PRT; 398 AA.
ID Q8NA87 HUMAN PRELIMINARY; PRT; 398 AA.
AC Q8NA87
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DE CDNA FLJ35741 fis, clone TEST12004163, moderately similar to Mus
DE musculus type II cytokeratin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Testis;
RX PubMed=14702039; DOI=10.1038/ngl285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hoshiro T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Mutsaers H., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihawa S., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya K., Momiya N., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
CC -I- SIMILARITY: Belongs to the intermediate filament family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
DR EMBL; AK093060; BAC04039.1; -; mRNA.
DR HSSP; P08670; 1GK7.
DR Ensembl; ENSG00000170486; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR1893:SP5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
DR Intermediate filament.
KW SEQUENCE 398 AA; 44698 MW; D5DD16028B8EBA3D CRC64;
SQ
Query Match 75.8%; Score 50; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 5.8;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLS 15
Db 344 SELNRLQIRIS 355
||||:|||||

RESULT 8
Q8WWZ0 HUMAN PRELIMINARY; PRT; 510 AA.
ID Q8WWZ0 HUMAN PRELIMINARY; PRT; 510 AA.
AC Q8WWZ0;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DE CDNA FLJ35741 fis, clone TEST12004163, moderately similar to Mus
DE musculus type II cytokeratin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1]
RN NUCLEOTIDE SEQUENCE.
RP Porter R.M., McLean W.H.I.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to the intermediate filament family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
DR EMBL; AY033495; AAK55108.1; -; mRNA.
DR HSSP; P08670; 1GK7.
DR Ensembl; ENSG00000170486; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR1893:SP5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
DR Intermediate filament; Keratin.
KW SEQUENCE 510 AA; 55749 MW; 93AAAF9D75378EE7 CRC64;
SQ
Query Match 75.8%; Score 50; DB 2; Length 510;
Best Local Similarity 83.3%; Pred. No. 5.8;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLS 15
Db 344 SELNRLQIRIS 355
||||:|||||

RESULT 9
Q8WWY9 HUMAN PRELIMINARY; PRT; 511 AA.
ID Q8WWY9 HUMAN PRELIMINARY; PRT; 511 AA.
AC Q8WWY9;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DE CDNA FLJ35741 fis, clone TEST12004163, moderately similar to Mus
DE musculus type II cytokeratin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1]
RN NUCLEOTIDE SEQUENCE.
RP Porter R.M., McLean W.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL
```

```

RP NUCLEOTIDE SEQUENCE.
RA McLean W.H.I.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY033496; AAK55109.1; -; Genomic_DNA.
DR HSPSP; P08670; ICK7.
DR Ensemble; ENSG00000170486; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR18993:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 511 AA; 55877 MW; 61E830C382063731 CRC64;

Query Match 75.8%; Score 50; DB 2; Length 511;
Best Local Similarity 83.3%; Pred. No. 5.8;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVQLRLRS 15
Db 344 SELNRQLRLRS 355
|||||:|||||

RESULT 10
K2CA_RAT
ID K2CA_RAT STANDARD; PRT; 536 AA.
AC QGIG00;
DT 22-NOV-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 14.
DE Keratin, type II cytoskeletal 4 (Cytokeratin-4) (CK-4) (Keratin-4)
DE (K4) (Type II keratin K04).
OS Rattus norvegicus (Rat).
GN Name=Krt4; Synonyms=K04;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Brown Norway.
RX PubMed=15057822; DOI=10.1038/nature02426;
RA Gibbs R.A., Weinstock G.M., Metzker M.L., Muzny D.M., Sodergren E.J.,
RA Scherer S., Scott G., Steffen D., Worley K.C., Burch P.E., Okwuonu G.,
RA Hines S., Lewis L., Deramo C., Delgado O., Dugan-Rocha S., Miner G.,
RA Morgan M., Hawes A., Gill R., Holt R.A., Adams M.D., Amanatides P.G.,
RA Baden-Tillson H., Barnstead M., Chin S., Evans C.A., Ferriera S.,
RA Poser C., Glodek A., Gu Z., Jennings D., Kraft C.L., Nguyen T.,
RA Franke K.C., Sitter C., Sutton G.G., Venter J.C., Woodage T.,
RA Smith D., Lee H.M., Gustafson E., Cahill P., Kana A.,
RA Doucette-Stamm L., Weinstock K., Fechtel K., Weiss R.B., Dunn D.M.,
RA Green E.D., Blakesley R.W., Bouffard G.G., De Jong P.J., Osogawa K.,
RA Zhu B., Marra M., Schein J., Bosdet I., Fjell C., Jones S.,
RA Krzywinski M., Mathewson C., Siddiqui A., Wye N., McPherson J.,
RA Zhao S., Fraser C.M., Shetty J., Shatsman S., Geer K., Chen Y.,
RA Abramson S., Nierman W.C., Havlak P.H., Chen R., Durbin K.J., Egan A.,
RA Ren Y., Song X.Z., Li B., Liu Y., Qin X., Cawley S., Cooney A.J.,
RA D'Souza L.M., Martin K., Wu J.O., Gonzalez-Garay M.L., Jackson A.R.,
RA Kalafus K.J., McLeod M.P., Milosavljevic A., Virk D., Volkov A.,
RA Wheeler D.A., Zhang Z., Bailey J.A., Eichler E.E., Tuzun E.,
RA Birney E., Mongin E., Ureta-Vidal A., Woodward C., Zdobnov E.,
RA Bork P., Suyama M., Torrents D., Alexandersson M., Trask B.J.,
RA Young J.M., Huang H., Wang H., Xing H., Daniels S., Gietzen D.,
RA Schmidt J., Stevens K., Vitt U., Wingrove J., Camara F., Mar Alba M.,
RA Abril J.F., Guigo R., Smit A., Dubchak I., Rubin E.M., Couronne O.,

```

```

RA Poliakov A., Hubner N., Ganten D., Goesele C., Hummel O., Kreitler T.,
RA Lee Y.A., Monti J., Schulz H., Zimdahl H., Himmelbauer H., Lehrach H.,
RA Jacob H.J., Bromberg S., Gullings-Handley J., Jensen-Seaman M.I.,
RA Kwitek A.E., Lazar J., Pasko D., Tonellato P.J., Twigger S.,
RA Ponting C.P., Duarte J.M., Rice S., Goodstadt L., Beaton S.A.,
RA Enes R.D., Winter E.E., Webber C., Brandt P., Nyakatura G.,
RA Adachi M., Charomonte F., Elnitski L., Eswara P., Hardison R.C.,
RA Hou M., Kolbe D., Makova K., Miller W., Nekrutenko A., Riemer C.,
RA Schwartz S., Taylor J., Yang S., Zhang Y., Lindpaintner K.,
RA Andrews T.D., Caccamo M., Clamp M., Clarke L., Curwen V., Durbin R.,
RA Eyras E., Seale S.M., Cooper G.M., Batzoglou S., Brudno M., Sidow A.,
RA Stone E.A., Payseur B.A., Bourque G., Lopez-Otin C., Puente X.S.,
RA Chakrabarti K., Chatterji S., Dewey C., Pachter L., Bray N., Yap V.B.,
RA Caspi A., Tesler G., Pevzner P.A., Haussler D., Roskin K.M.,
RA Baertsch R., Clawson H., Purrey T.S., Hinrichs A.S., Karolchik D.,
RA Kent W.J., Rosenbloom K.R., Trumbower H., Weirauch M., Cooper D.N.,
RA Stenson P.D., Ma B., Brent M., Arumugam M., Shteynberg D.,
RA Copley R.R., Taylor M.S., Rietman H., Mudunuri U., Peterson J.,
RA Guyer M., Felsenfeld A., Old S., Mockrin S., Collins F.,
RT "Genome sequence of the Brown Norway rat yields insights into
RT mammalian evolution.";
RL Nature 428:493-521 (2004).
RN [2]
RN IDENTIFICATION.
RP MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;
RX Hesse M., Zimek A., Weber K., Magin T.M.;
RA "Comprehensive analysis of keratin gene clusters in humans and
RT rodents.";
RL Eur. J. Cell Biol. 83:19-26(2004).
CC -1- SUBUNIT: Heterotetramer of two type I and two type II keratins.
CC Keratin 4 is generally associated with keratin 13.
CC -1- MISCELLANEOUS: There are two types of cytoskeletal and
CC microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II
CC (neutral to basic; 56-70 kDa) [K1 to K8].
CC -1- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AABR03057015; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR EMBL; BK003985; DAA02230.1; -; mRNA.
DR Ensemble; ENSRNCG00000032332; Rattus norvegicus.
DR RGD; 1359272; K04.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR18993:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament; Keratin; Phosphorylation.
FT Keratin, type II cytoskeletal 4.
FT /FTid=PRO_0000063724.
CHAIN 1 536
REGION 1 145 Head.
REGION 146 454 Rod.
REGION 146 181 Coil 1A.
REGION 182 200 Linker 1.
REGION 201 293 Coil 1B.
REGION 294 316 Linker 12.
REGION 317 454 Coil 2.
REGION 455 524 Tail.
REGION 525 511 Gly-rich.
FT COMPTIAS 471 511 Gly-rich.
FT SITE 395 395 Stutter.
FT MOD RES 51 51 Phosphoserine (By similarity).
SQ SEQUENCE 536 AA; 57667 MW; 5D1AC47FA0477DB4 CRC64;

Query Match 75.8%; Score 50; DB 1; Length 536;
Best Local Similarity 83.3%; Pred. No. 6.1;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVQLRLRS 15

```

DB 365 SELNRMQIRS 376  
|||||:||||:  
RESULT 11  
Q8BG27 MOUSE PRELIMINARY; PRT; 551 AA.  
AC Q8BG27  
DC  
DT 01-WAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-WAR-2006, sequence version 1.  
DE 07-FEB-2006, entry version 21.  
DE 10 days neonate skin cDNA, RIKEN full-length enriched library,  
DE clone:4732475103 product:CYTOKERATIN homolog (6 days neonate skin  
DE cDNA, RIKEN full-length enriched library, clone:A030012D14  
DE product:CYTOKERATIN homolog) (10 days neonate skin cDNA, RIKEN full-  
DE length enriched library, clone:4732468K03 product:CYTOKERATIN TYPE II  
DE homolog).  
GN Names:Krtcap1;  
GN Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Belsel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.D., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Furuki S., Gariboldi M.,  
RA Georgii-Heminger P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminek L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nieuwen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashina T., Kojima M., Kondo S., Konno H., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563 (2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RT (Genome Network Core Team) and the FANTOM Consortium;  
RL "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Otsu N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarini R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKensie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wymshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ikeni Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yashinashi E., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Bash G.,  
RA Blake J., Offelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Schor K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wymshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Science 309:1559-1563 (2005).





RT bullosa simplex.";  
RL Mol. Biol. Cell 12:1775-1789(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=Czech II; TISSUE=Mammary gland;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
CC Keratin-5 associates with keratin-14.  
CC -1- TISSUE SPECIFICITY: Expressed in epidermis.  
CC -1- MISCELLANEOUS: There are two types of cytoskeletal and  
CC microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II  
CC (neutral to basic; 56-70 kDa) [K1 to K8].  
CC -1- SIMILARITY: Belongs to the intermediate filament family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR ENBL; AF306785; AAL16774.1; -; Genomic\_DNA.  
DR ENBL; BC006780; AAH06780.1; -; mRNA.  
DR HSSP; P08670; 1GK7.  
DR IntAct; Q922U2; -.  
DR Ensembl; ENSMUSG00000061527; Mus musculus.  
DR MGI; MGI:196702; Krt2-5.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR003054; Keratin\_II.  
DR PANTHER; PTHR18893:SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Coiled coil; Intermediate filament; Keratin; Phosphorylation.  
FT CHAIN 1 580 Keratin, type II cytoskeletal 5.  
FT /FTID=PRO\_0000063728.  
FT REGION 1 161 Head.  
FT REGION 162 471 Rod.  
FT REGION 162 197 Coiled coil 1A.  
FT REGION 198 216 Linker 1.  
FT REGION 217 309 Coiled coil 1B.  
FT REGION 310 332 Linker 12.  
FT REGION 333 471 Coiled coil 2.  
FT REGION 472 580 Tail.  
FT COMPTAS 2 38 Ser-rich.  
FT COMPTAS 39 133 Gly-rich.  
FT COMPTAS 495 562 Gly-rich.  
FT SITE 413 413 Stutter.  
FT MOD RES 61 61 Phosphoserine (By similarity).  
FT CONFLICT 139 139 V -> I (in Ref. 1).  
SQ SEQUENCE 580 AA; 61767 MW; 304B341FD4224B41 CRC64;  
  
Query Match 75.8%; Score 50; DB 1; Length 580;  
Best Local Similarity 83.3%; Pred. No. 6.6;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 4 SELNRVQLRLS 15

Db 381 SEMNRMIQRLRS 392

Search completed: July 5, 2006, 19:01:35  
Job time : 164.828 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 139.029 Seconds

(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928a-9

Perfect score: 66

Sequence: 1 XXYSELNRVQLRLSI 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq\_8:\*

1: geneseqp1980s:\*

2: geneseqp1990s:\*

3: geneseqp2000s:\*

4: geneseqp2001s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

9: geneseqp2005s:\*

10: geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	66	100.0	16	6 AAO26395	Aao26395 Psoriasis
2	66	100.0	16	9 ADV04414	Adv04414 Leishmania
3	59	89.4	16	6 AAO26389	Aao26389 Psoriasis
4	55	83.3	643	6 ABG74869	Abg74869 Human cyt
5	55	83.3	644	8 ADO17549	Ado17549 Human sof
6	55	83.3	644	9 AEA15501	Aea15501 Human pol
7	52	78.8	628	8 ADS85115	Ads85115 Mouse ato
8	50	75.8	398	7 ADB65234	Adb65234 Human pro
9	50	75.8	511	5 ABG94648	Abg94648 Human NOV
10	50	75.8	511	7 ADL36000	Adl36000 Human NOV
11	50	75.8	511	9 AEA15498	Aea15498 Human pol
12	50	75.8	601	8 ADM07163	Adm07163 Bos tauru
13	50	75.8	679	9 AEA20243	Aea20243 Novel hum
14	47	71.2	20	9 AED42282	Aed42282 Human pro
15	47	71.2	92	8 ADR98933	Adr98933 Lung spec
16	47	71.2	268	8 ADR98973	Adr98973 Lung spec
17	47	71.2	295	8 ADR98974	Adr98974 Lung spec
18	47	71.2	342	7 ADE15641	Ade15641 Human str
19	47	71.2	551	3 AAY52398	Aay52398 Human ker
20	47	71.2	551	5 AAE20423	Aae20423 Human ker
21	47	71.2	564	5 ABP65226	Abp65226 Hypoxia-r
22	47	71.2	564	8 ADJ75686	Adj75686 Marker ge
23	47	71.2	564	8 ADN04684	Adn04684 Antipsori

24	47	71.2	564	8 ADN04384	Adn04384 Antipsori
25	47	71.2	564	8 ADO55108	Ado55108 Protein #
26	47	71.2	564	8 ADQ18800	Adq18800 Human sof
27	47	71.2	564	9 AEA15493	Aea15493 Human pol
28	47	71.2	564	9 AEA15491	Aea15491 Human pol
29	47	71.2	564	9 AEA15492	Aea15492 Human pol
30	47	71.2	564	9 AEA15494	Aea15494 Human pol
31	47	71.2	564	9 AEB17592	Aeb17592 Human ker
32	47	71.2	564	9 AEB17594	Aeb17594 Human ker
33	47	71.2	564	9 AEB87757	Aeb87757 Human cyt
34	47	71.2	567	7 ADB65618	Adb65618 Human pro
35	47	71.2	585	9 AED74047	Aed74047 Human pla
36	47	71.2	602	6 ABR41632	Abr41632 Human DIT
37	47	71.2	625	6 ABP98825	Abp98825 Human str
38	47	71.2	645	5 ABG77177	Abg77177 Prostata
39	47	71.2	645	8 ADN04309	Adn04309 Antipsori
40	47	71.2	645	9 ADV85790	Adv85790 Human pro
41	47	71.2	645	9 AEA15500	Aea15500 Human pol
42	46	69.7	327	7 ADJ69669	Adj69669 Human hea
43	46	69.7	508	8 ADS10914	Ads10914 Human hea
44	46	69.7	521	5 ABG94649	Abg94649 Human NOV
45	46	69.7	521	5 ABB78802	Abb78802 Human NOV

#### ALIGNMENTS

RESULT 1  
AAO26395  
ID AAO26395 standard; peptide; 16 AA.  
AC AAO26395;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
PA (AKIV-) AKIVA LLC.  
XX  
PI O'daly JA;  
XX  
DR WPI; 2003-018763/01.  
XX  
PT New polypeptide and immunogenic variants comprising amino acid sequences  
PT of particulate antigens, useful for the treatment and clinical remission  
PT of psoriasis.  
XX  
PS Example 17; Page 43; 56pp; English.  
XX  
CC The invention relates to a polypeptide comprising an isolated amino acid  
CC sequence or immunogenic variants selected from any of 14 fully defined  
CC sequences of 7-16 amino acids, given in the specification. The  
CC immunotherapeutic agents and a pharmaceutical compositions comprising  
CC polynucleotides and vectors of the invention are useful for the treatment  
CC and clinical remission of psoriasis. The isolated nucleic acids are  
CC useful as probes. The sequences of the invention can be used in the  
CC treatment of disorders by gene therapy. This sequence represents an  
CC immunotherapeutic peptide related to the invention  
XX

SQ Sequence 16 AA;

Query Match 100.0%; Score 66; DB 6; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 0.00017;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YSELNRVQLRLSI 16  
 |||||  
 Db 3 YSELNRVQLRLSI 16

RESULT 2  
 ADV04414  
 ID ADV04414 standard; peptide; 16 AA.  
 XX  
 AC ADV04414;  
 XX  
 DT 24-FEB-2005 (first entry)  
 XX  
 DE Leishmania immunotherapeutic peptide SEQ ID NO:9.  
 XX  
 KW T-lymphocyte; immunogenicity; antigen; antiprosoriatic; immunostimulant;  
 XX  
 KW psoriasis.  
 XX  
 OS Leishmania sp.  
 XX  
 PN US2004241168-A1.  
 XX  
 PD 02-DEC-2004.  
 XX  
 PF 09-FEB-2004; 2004US-00774928.  
 XX  
 PR 16-MAR-2001; 2001US-00809003.  
 PR 17-OCT-2003; 2003US-00687892.  
 XX  
 PA (ODAL/) ODALY J A.  
 XX  
 PI Odaly JA;  
 XX  
 XX WPI; 2005-011563/01.  
 XX  
 XX Inhibiting selectively T-cell rolling in human, by administering compound  
 PT interfering with cutaneous lymphocyte antigen-E selectin interaction and  
 PT leukocyte function associated antigen-1 interaction.  
 XX  
 PS Claim 18; SEQ ID NO 9; 21pp; English.  
 XX  
 CC The invention relates to a novel method for selectively inhibiting T-cell  
 CC rolling in a human host, comprising administering a compound that  
 CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
 CC selectin interaction and leukocyte function associated antigen (LFA)-  
 CC 1/intercellular adhesion molecule (ICAM) and very late antigen  
 CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
 CC includes an immunotherapeutic agent, which comprises a purified protein  
 CC extract that is isolated by diethylaminoethyl Sephadex chromatography of  
 CC a Nonidet P-40 insoluble particulate antigen fraction derived from  
 CC isolated killed cells of amastigotes from one or more species of the  
 CC Leishmania genus, where the particulate antigen fraction is solubilized  
 CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
 CC to diethylaminoethyl Sephadex and is eluted with a solution comprising  
 CC 0.1 M sodium chloride, 8 M urea and 0.025 M  
 CC Tris(hydroxymethyl)aminomethane pH 8.3. and the purified protein extract  
 CC includes polypeptides having apparent molecular weights of 73, 80 and 82  
 CC kDa, after total reduction and alkylation. The species is L. amazonensis,  
 CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the  
 CC invention has antiprosoriatic and immunostimulant activity, and inhibits T-  
 CC cell rolling by interfering with CLA-E selectin interaction and LFA-  
 CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
 CC inhibiting T-cell rolling in a human host, and for treating psoriasis.  
 CC The present sequence represents a Leishmania peptide used in the  
 CC invention.  
 XX  
 XX Sequence 16 AA;

Query Match 100.0%; Score 66; DB 9; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 0.00017;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YSELNRVQLRLSI 16  
 |||||  
 Db 3 YSELNRVQLRLSI 16

RESULT 3  
 AAO26389  
 ID AAO26389 standard; peptide; 16 AA.  
 XX  
 AC AAO26389;  
 XX  
 DT 30-JAN-2003 (first entry)  
 XX  
 DE Psoriasis treating immunotherapeutic peptide, SEQ ID NO 9.  
 XX  
 KW Antiprosoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
 KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
 KW gene therapy.  
 XX  
 OS Leishmania sp.  
 XX  
 PN WO200274239-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 04-MAR-2002; 2002WO-US006496.  
 XX  
 PR 16-MAR-2001; 2001US-00809003.  
 XX  
 PA (AKIV-) AKIVA LLC.  
 XX  
 PI O'daly JA;  
 XX  
 XX WPI; 2003-018763/01.  
 XX  
 XX New polypeptide and immunogenic variants comprising amino acid sequences  
 PT of particulate antigens, useful for the treatment and clinical remission  
 PT of psoriasis.  
 XX  
 PS Claim 1; Page 55; 56pp; English.  
 XX  
 CC The invention relates to a polypeptide comprising an isolated amino acid  
 CC sequence or immunogenic variants selected from any of 14 fully defined  
 CC sequences of 7-16 amino acids, given in the specification. The  
 CC immunotherapeutic agents and a pharmaceutical compositions comprising  
 CC polynucleotides and vectors of the invention are useful for the treatment  
 CC and clinical remission of psoriasis. The isolated nucleic acids are  
 CC useful as probes. The sequences of the invention can be used in the  
 CC treatment of disorders by gene therapy. This sequence represents one of  
 CC the 14 immunotherapeutic peptides of the invention  
 XX  
 XX Sequence 16 AA;

Query Match 89.4%; Score 59; DB 6; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 0.0029;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLSI 16  
 |||||  
 Db 4 SELNRVQLRLSI 16

RESULT 4  
 ABG74869  
 ID ABG74869 standard; protein; 643 AA.  
 XX  
 AC ABG74869;  
 XX

DT 19-JUN-2003 (first entry)  
XX Human cytokерatin-1 protein.  
DE Cytokeratin-1; inflammation; infection; soluble; sepsis; veterinary;  
XX sepsis-like systemic infection; human.  
KW Homo sapiens.  
XX WO2003002600-A1.  
PN 09-JAN-2003.  
XX 12-JUN-2002; 2002WO-EP006473.  
PF 27-JUN-2001; 2001DE-01030985.  
XX (BRAH-) BRAHMS AG.  
PA Bergmann A, Struck J, Uehlein M;  
PI WPI; 2003-201491/19.  
XX New use of soluble cytokерatin-1 fragments from body fluid are tissue as  
PT marker peptides or to provide therapeutics for the diagnosis, prognosis  
PT and treatment of inflammatory disease and infection, including systemic  
PT inflammation.  
XX Claim 3; Page 38-40; 51pp; German.  
XX This invention describes a novel use of soluble cytokерatin-1 fragments  
CC from body fluid or tissue as marker peptides for the detection, prognosis  
CC and control of inflammation and infection or as a therapeutic target for  
CC inflammation and infection. The invention also describes a method of  
CC producing a soluble cytokерatin-1 fragment in rich or pure form from a  
CC human material rich in cytokерatin-1 by enzymatically splitting the  
CC cytokерatin-1 with endoprotease to obtain the hydrolysis products and  
CC purifying the resulting fragments. The novel cytokерatin fragments are  
CC useful for the differential diagnosis, early recognition and prognosis,  
CC evaluation of severity of sepsis and heavy infection, particularly sepsis  
CC -like systemic infection. The fragments are used in human and veterinary  
CC therapy, particularly the fragment or it's specific antibody is used to  
CC treat inflammatory disease or infection. This sequence represents the  
CC human cytokерatin-1 protein described in the disclosure of the invention  
XX  
SQ Sequence 643 AA;  
Query Match 83.3%; Score 55; DB 6; Length 643;  
Best Local Similarity 100.0%; Pred. No. 0.83;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 SELNRVQLRLS 15  
Db |||||  
398 SELNRVQLRLS 409  
RESULT 5  
ADQL7549  
ID ADQL7549 standard; protein; 644 AA.  
XX AC ADQL7549;  
XX 26-AUG-2004 (first entry)  
XX Human soft tissue sarcoma-upregulated protein - SEQ ID 366.  
DE soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
XX Homo sapiens.  
OS WO2004048938-A2.  
PN 10-JUN-2004.  
PD

XX 26-NOV-2003; 2003WO-US038193.  
XX 26-NOV-2002; 2002US-0429739P.  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX Aziz N, Ginsburg WM, Zlotnik A;  
XX WPI; 2004-441208/41.  
XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX Example 2; SEQ ID NO 366; 210pp; English.  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cyostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC protein of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX Sequence 644 AA;  
SQ Query Match 83.3%; Score 55; DB 8; Length 644;  
Best Local Similarity 100.0%; Pred. No. 0.83;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 SELNRVQLRLS 15  
Db |||||  
399 SELNRVQLRLS 410  
RESULT 6  
AEAL5501  
ID AEAL5501 standard; protein; 644 AA.  
XX AC AEAL5501;  
XX 28-JUL-2005 (first entry)  
XX Human polypeptide #120.  
DE Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
XX colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
XX cytostatic; neoplasm.  
XX Homo sapiens.  
OS WO2005047534-A2.  
PN 26-MAY-2005.  
XX 15-OCT-2004; 2004WO-EP011599.  
XX 28-OCT-2003; 2003EP-00024565.  
XX (FARB ) BAYER HEALTHCARE AG.  
PA Wirtz R, Munnes M;  
PI WPI; 2005-372393/38.  
XX N-PSDB; AEAL5423.  
DR

```

XX Predicting a response to cancer treatment by detecting at least 2
PT markers, which are genes or genomic nucleic acid sequences that are
PT located on one chromosomal region, which is altered in malignant
PT neoplasia.
XX
XX Claim 7; SEQ ID NO 463; 464pp; English.
XX
CC The invention relates to a method of predicting response to cancer
CC treatment comprising detection of at least 2 markers, where the markers
CC are genes and fragments or genomic nucleic acid sequences that are
CC located on one chromosomal region, which is altered in malignant
CC neoplasia. The invention also relates to a method for the prediction,
CC diagnosis or prognosis of malignant neoplasia, methods for detecting
CC deregulations in malignant neoplasia and breast cancer, a method of
CC determining the phenotype of a cell or tissue, a method for identifying
CC genomic regions which are altered on the chromosomal level and encode
CC genes that are linked by function and are differentially expressed in
CC malignant neoplasia and breast cancer, methods of screening for agents
CC which regulate the activity of a polypeptide or a polynucleotide and
CC antibodies that specifically bind to a full length or partial
CC polypeptide. The method is useful for predicting response to cancer
CC treatment. The methods and compositions are useful for predicting,
CC diagnosing, prognosing, preventing or treating malignant neoplasia
CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,
CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell
CC lung cancer. This sequence represents a human polypeptide used in the
CC scope of the invention.
XX
SQ Sequence 644 AA;
Query Match 83.3%; Score 55; DB 9; Length 644;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRLS 15
DB 399 SELNRVIOQLRLS 410
|||||
|||||

RESULT 7
AD8585115
ID ADS85115 standard; protein; 628 AA.
XX
AC ADS85115;
XX
XX 19-NOV-2004 (first entry)
XX
DE Mouse atopic dermatitis-related protein sequence SeqID117.
XX
XX atopic dermatitis; gene expression level; skin; inflammation; rash;
KW Dermatological; antiinflammatory; antipsoriatic; psoriasis; mouse;
KW murine.
XX
XX Mus musculus.
XX
XX WQ2004031386-A1.
XX
PD 15-APR-2004.
XX
XX 01-AUG-2003; 2003WO-JP009808.
XX
PF 06-AUG-2002; 2002JP-00229318.
PR 14-MAY-2003; 2003JP-00136543.
XX
XX (GENO-) GENOX RES INC.
PA (UYJU-) UNIV JUNTENDO.
XX
XX Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;
PI Mitsuishi K;
XX
XX WPI; 2004-330185/30.
DR N-PSDB; ADS85114.
XX

```

---

```

XX Determination of difference in expression level of specified genes in
PT inflamed and non-inflamed skin sites for diagnosis and examination of
PT atopic dermatitis and psoriasis.
XX
XX Example 1; SEQ ID NO 117; 611pp; Japanese.
XX
CC This invention relates to a novel method for the examination of atopic
CC dermatitis in which the expression level of specified genes in specimens
CC of skin taken from inflamed (rash) areas and non-inflamed areas is
CC compared and the presence of increased or reduced expression in the
CC inflamed areas determined. The invention may be useful for the
CC development of compounds with a dermatological, antiinflammatory or
CC antipsoriatic activity acting as inhibitors and stimulators of genes
CC involved in atopic dermatitis and psoriasis. The invention may be useful
CC for treatment, prevention, diagnosis and assessment of atopic dermatitis
CC and psoriasis. The present sequence is that of an atopic dermatitis-
CC related protein which was used in the exemplification of the invention.
XX
SQ Sequence 628 AA;
Query Match 78.8%; Score 52; DB 8; Length 628;
Best Local Similarity 91.7%; Pred. No. 2.7;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRLS 15
DB 407 SELNRVIOQLRLS 418
|||||
|||||

RESULT 8
ADB65234
ID ADB65234 standard; protein; 398 AA.
XX
AC ADB65234;
XX
XX 04-DEC-2003 (first entry)
XX
DE Human protein encoded by clone TESTI20041630.
XX
XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW cell regeneration; membrane protein; signal transduction-related protein;
KW transcription-related protein; osteoporosis; neurological disease;
KW cancer; tumour.
XX
XX Homo sapiens.
XX
XX EP1308459-A2.
XX
XX 07-MAY-2003.
XX
XX 28-MAR-2002; 2002EP-00007401.
XX
XX 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI; 2003-450961/43.
DR N-PSDB; ADB63264.
XX
XX New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
XX Claim 1; Page; 222pp; English.
XX
XX The invention discloses a polynucleotide comprising a sequence selected

```

CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours). The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a protein of the invention. Note: Some of the  
 CC sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.

XX SQ Sequence 398 AA;

Query Match 75.8%; Score 50; DB 7; Length 398;  
 Best Local Similarity 83.3%; Pred. No. 3.7;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVLIQRLRS 15  
 |||||:||||:  
 Db 231 SELNRLLIQRIRS 242

RESULT 9  
 ABG94648  
 ID ABG94648 standard; protein; 511 AA.

XX AC ABG94648;

XX DT 02-DEC-2002 (first entry)

XX DE Human NOV10b protein.

XX Human; NOVX; pathological condition; NOVX-associated disorder; diabetes;  
 KW Von Hippel-Lindau syndrome; cirrhosis; transplantation disorder; obesity;  
 KW pancreatitis; autoimmune disease; renal artery stenosis; infertility;  
 KW interstitial nephritis; glomerulonephritis; polycystic kidney disease;  
 KW systemic lupus erythematosus; SLE; cataract; Alzheimer's disease;  
 KW acoustic trauma; cancer; cardiomyopathy; atherosclerosis; hypertension;  
 KW congenital heart defect; scleroderma; endometriosis; haemophilia;  
 KW dementia; stroke; Parkinson's disease; Huntington's disease; epilepsy;  
 KW multiple sclerosis; anxiety; pain; leukaemia; hypothyroidism; psoriasis;  
 KW acne; wound; asthma.

XX OS Homo sapiens.

XX PN WO200266643-A2.

XX PD 29-AUG-2002.

XX PF 13-NOV-2001; 2001WO-US048732.

XX PR 13-NOV-2000; 2000US-0248153P.

XX PR 17-NOV-2000; 2000US-0249598P.

XX PR 26-JAN-2001; 2001US-0264240P.

XX PR 02-FEB-2001; 2001US-0266127P.

XX PR 16-FEB-2001; 2001US-0269562P.

XX PR 10-JUL-2001; 2001US-0304348P.

XX PR 31-JUL-2001; 2001US-0309261P.

XX PR 17-AUG-2001; 2001US-0313283P.

PA (CURA-) CURAGEN CORP.

XX Malyankar UM, Shenoy SG, Spytek KA, Zerhusen BD, Patturajan M;  
 PI Guo X, Kekuda R, Gangolli EA, Shimkets RA, Taupier RJ, Li L;  
 PI Padigaru M;

XX WPI; 2002-706943/76.

XX DR N-PSDB; ABS71710.

XX New isolated NOVX polypeptides and nucleic acid molecules useful for  
 PT treating, preventing, diagnosing and researching of pathological  
 PT conditions in humans with a NOVX-associated disorders.

XX Claim 1; Page 112; 295pp; English.

XX The present invention relates to new NOVX polypeptides. The NOVX  
 CC polypeptide, nucleic acid and antibody are useful for treating or  
 CC preventing a pathological condition in humans with a NOVX-associated  
 CC disorder, e.g. Von Hippel-Lindau syndrome, cirrhosis, transplantation  
 CC disorders, pancreatitis, obesity, diabetes, autoimmune disease, renal  
 CC artery stenosis, interstitial nephritis, glomerulonephritis, polycystic  
 CC kidney disease, systemic lupus erythematosus (SLE), cataract, Alzheimer's  
 CC disease, acoustic trauma, cancer, infertility, cardiomyopathies,  
 CC atherosclerosis, haemophilia, dementia, stroke, Parkinson's disease,  
 CC endometriosis, epilepsy, multiple sclerosis, anxiety, pain,  
 CC Huntington's disease, hypothyroidism, psoriasis, acne, wounds and asthma. They are  
 CC also useful for the manufacture of a medicament for treating a syndrome  
 CC associated with a human disease, specifically a NOVX-associated disorder.  
 CC They may also be useful in therapeutic applications including protein  
 CC therapeutic, small molecule drug target, antibody target, diagnostic  
 CC and/or prognostic marker, gene therapy, research tools and tissue  
 CC regeneration. The present amino acid sequence represents a NOVX protein  
 CC of the invention

XX SQ Sequence 511 AA;

Query Match 75.8%; Score 50; DB 5; Length 511;

Best Local Similarity 83.3%; Pred. No. 4.8;

Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVLIQRLRS 15

|||||:||||:  
 Db 344 SELNRLLIQRIRS 355

RESULT 10

ADL36000

ID ADL36000 standard; protein; 511 AA.

XX AC ADL36000;

XX DT 20-MAY-2004 (first entry)

XX DE Human NOVX polypeptide #23.

XX Human; NOVX; Alzheimer's disease; Parkinson's disease; stroke; epilepsy;  
 KW multiple sclerosis; addiction; anxiety; pain; acne; alopecia;  
 KW inflammation; rheumatoid arthritis; AIDS; cancer; psoriasis;  
 KW hypertension; renal disorder; bone disease; haematopoietic disorder;  
 KW wound; bacterial infection; viral infection; fungal infection;  
 KW protozoal infection; urinary retention; osteoporosis;  
 KW myocardial infarction; diabetes; ulcer; cirrhosis; depression.

XX OS Homo sapiens.

XX PN US2003207800-A1.

XX PD 06-NOV-2003.

XX PF 13-NOV-2001; 2001US-00015115.

XX PR 08-NOV-1999; 99US-0164240P.

PR 13-NOV-2000; 2000US-0248153P.  
 PR 17-NOV-2000; 2000US-0249598P.  
 PR 02-FEB-2001; 2001US-0266127P.  
 PR 16-FEB-2001; 2001US-0269562P.  
 PR 10-JUL-2001; 2001US-0304348P.  
 PR 31-JUL-2001; 2001US-0309261P.  
 PR 17-AUG-2001; 2001US-0313283P.  
 XX (MALY/) MALYANKAR U M.  
 PA (SHEN/) SHENOY S G.  
 PA (SPYT/) SPYTEK K A.  
 PA (ZERH/) ZERHUSEN B D.  
 PA (PATT/) PATTURAJAN M.  
 PA (GUOX/) GUO X.  
 PA (KEKU/) KEKUDA R.  
 PA (GANG/) GANGOLLI E A.  
 PA (SHIM/) SHIMKETS R A.  
 PA (TAUP/) TAUPIER R J.  
 PA (LILL/) LI L.  
 PA (PADI/) PADIGARU M.  
 XX Malyankar UM, Shenoy SG, Spytek KA, Zerhusen BD, Patturajan M;  
 PI Guo X, Kekuda R, Gangolli EA, Shimkets RA, Taupier RJ, Li L;  
 PI Padigar M;  
 XX WPI: 2003-875894/81.  
 DR N-PSDB; ADJ35999.  
 XX New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOVX-associated disorders (e.g. stroke, epilepsy,  
 PT AIDS, pain, diabetes or cancer) and in chromosome mapping, tissue typing  
 PT or pharmacogenomics.  
 XX Claim 1; SEQ ID NO 46; 233pp; English.  
 XX The invention relates to human NOVX polypeptides and the polynucleotides  
 CC encoding them. The invention also relates to antibodies that bind  
 CC immunospecifically to the polypeptides. The NOVX polypeptides  
 CC polynucleotides and antibodies are useful in diagnosing, treating or  
 CC preventing NOVX-associated disorders such as Alzheimer's disease,  
 CC Parkinson's disease, stroke, epilepsy, multiple sclerosis, addiction,  
 CC anxiety, pain, acne, alopecia, inflammation, rheumatoid arthritis, AIDS,  
 CC cancer, psoriasis, hypertension, renal disorders, bone diseases,  
 CC haematopoietic disorders, wounds, infection (e.g. bacterial, viral,  
 CC fungal or protozoal), urinary retention, osteoporosis, myocardial  
 CC infarction, diabetes, ulcer, cirrhosis or depression. The polypeptides  
 CC are also useful as vaccines. This sequence represents a human NOVX  
 CC polypeptide of the invention.  
 XX SQ Sequence 511 AA;  
 Query Match 75.8%; Score 50; DB 7; Length 511;  
 Best Local Similarity 83.3%; Pred. No. 4.8;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 SELNRVLIQRLRS 15  
 Db |||||:||||:|  
 344 SELNRLIQRLRS 355  
 RESULT 11  
 AEA15498  
 ID AEA15498 standard; protein; 511 AA.  
 XX AEA15498;  
 AC AEA15498;  
 XX 28-JUL-2005 (first entry)  
 DT Human polypeptide #117.  
 XX Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
 KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
 KW cytotatic; neoplasm.

XX Homo sapiens.  
 OS WO2005047534-A2.  
 PN 26-MAY-2005.  
 PD 15-OCT-2004; 2004WO-EP011599.  
 PF 28-OCT-2003; 2003EP-00024565.  
 PR (FARB ) BAYER HEALTHCARE AG.  
 PA Wirtz R, Munnes M;  
 PI WPI: 2005-372393/38.  
 DR N-PSDB; AEA15420.  
 XX Predicting a response to cancer treatment by detecting at least 2  
 PT markers, which are genes or genomic nucleic acid sequences that are  
 PT located on one chromosomal region, which is altered in malignant  
 PT neoplasia.  
 XX Claim 7; SEQ ID NO 460; 464pp; English.  
 XX The invention relates to a method of predicting response to cancer  
 CC treatment comprising detection of at least 2 markers, where the markers  
 CC are genes and fragments or genomic nucleic acid sequences that are  
 CC located on one chromosomal region, which is altered in malignant  
 CC neoplasia. The invention also relates to a method for the prediction,  
 CC diagnosis or prognosis of malignant neoplasia, methods for detecting  
 CC deregulations in malignant neoplasia and breast cancer, a method of  
 CC determining the phenotype of a cell or tissue, a method for identifying  
 CC genomic regions which are altered on the chromosomal level and encode  
 CC genes that are linked by function and are differentially expressed in  
 CC malignant neoplasia and breast cancer, methods of screening for agents  
 CC which regulate the activity of a polypeptide or a polynucleotide and  
 CC antibodies that specifically bind to a full length or partial  
 CC polypeptide. The method is useful for predicting response to cancer  
 CC treatment. The methods and compositions are useful for predicting,  
 CC diagnosing, prognosing, preventing or treating malignant neoplasia  
 CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
 CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
 CC lung cancer. This sequence represents a human polypeptide used in the  
 CC scope of the invention.  
 XX SQ Sequence 511 AA;  
 Query Match 75.8%; Score 50; DB 9; Length 511;  
 Best Local Similarity 83.3%; Pred. No. 4.8;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 SELNRVLIQRLRS 15  
 Db |||||:||||:|  
 344 SELNRLIQRLRS 355  
 RESULT 12  
 ADW07163  
 ID ADW07163 standard; protein; 601 AA.  
 XX ADW07163;  
 AC ADW07163;  
 DT 10-MAR-2005 (first entry)  
 XX Bos taurus keratin 5 protein sequence SeqID3.  
 DE animal breeding; DNA type; epidermolysis bullosa simplex; keratin.  
 XX Bos taurus.  
 OS Bos taurus.  
 XX Key Location/Qualifiers  
 FH Misc-difference 478  
 FT

/note= "wild-type glutamic acid may be substituted by lysine as the result of a single nucleotide polymorphism"

```

FT  /note= "wild-type glutamic acid may be substituted by
XX  lysine as the result of a single nucleotide polymorphism"
PN  NZ525023-A.
XX
XX  27-AUG-2004.
XX
XX  28-MAR-2003; 2003NZ-00525023.
XX
XX  28-MAR-2003; 2003NZ-00525023.
XX
XX  (SNEL/) SNELL R G.
PA  (FORD/) FORD C A.
PA  (SUTH/) SUTHERLAND G T.
PA  (STAN/) STANFIELD A M.
PA  (SPEL/) SPELMAN R J.
PA  (ANKE/) ANKERSMIT-UDY A E L.
XX
XX  Snell RG, Ford CA, Sutherland GT, Stanfield AM, Spelman RJ;
PI  Ankersmit-Udy AEL;
XX
XX  WPI; 2004-811782/80.
DR  N-PSDB; ADW07161, ADW07162.
XX
XX  Assessing genetic merit of bovine with respect to epidermolysis bullosa
PT  simplex involves determining keratin 5 genotypic state of bovine.
XX
XX  Claim 4; SEQ ID NO 3; 40pp; English.
PS
XX
XX  This invention relates to a novel method for assessing the genetic merit
CC  of a bovine with respect to epidermolysis bullosa (EB) simplex which
CC  involves determining the keratin 5 genotypic state of the bovine. The
CC  method of the invention enables a farmer or animal breeder to eliminate
CC  or select against the disease (EB) in their animal population. The
CC  present sequence is that of the bovine keratin 5 protein and which was
CC  used in the method of the invention.
XX
XX  Sequence 601 AA;
SQ
XX
Query Match 75.8%; Score 50; DB 8; Length 601;
Best Local Similarity 83.3%; Pred. No. 5.8;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  4 SELNRVIOQLRS 15
DB  388 SEMNRMIQRLRS 399
    ||:|||||
    ||:|||||

RESULT 13
AEA20243
ID  AEA20243 standard; protein; 679 AA.
XX
XX  AEA20243;
AC
XX
XX  11-AUG-2005 (first entry)
DT
XX
DE  Novel human polypeptide SEQ ID NO 937.
XX
XX  vulnery; CNS-gen.; gene therapy; diagnostic; forensic; mapping;
KW  DNA purification; protein purification; osteoarthritis; antiarthritic;
KW  osteopathic; musculoskeletal disease; osteoporosis; endocrine disease;
KW  periodontal disease; antiinflammatory; mouth disease; burns; injury;
KW  peripheral neuropathy; Alzheimer's disease; neuroprotective; neurologic;
KW  degeneration; parkinson's disease; antiparkinsonian; neurological disease;
KW  cerebrovascular ischemia; cerebroprotective; vasotropic;
KW  cardiovascular disease; autoimmune disease; immunosuppressive;
KW  immune disorder; viral infection; virucide; infection; cancer;
KW  cytostatic; neoplasm.
XX
XX  Homo sapiens.
OS
XX
XX  WO2005049806-A2.
PN
XX

```

```

PD  02-JUN-2005.
XX
XX  11-MAR-2004; 2004WO-US007412.
PF
XX
XX  14-MAR-2003; 2003US-00389559.
PR
XX
XX  (NUVE-) NUVELO INC.
PA
XX
XX  Tang TY, Wang J, Wang ZW, Zhang J, Ren F, Zhou P, Ma Y;
PI  Ghosh M, Xue A, Asundi V, Zhao Q, Wang D, Goodrich R, Chen R;
PI  Wehrman T, Weng G, Boyle B;
XX
XX  WPI; 2005-417730/42.
DR  N-PSDB; AEA19676.
XX
XX  New polynucleotide encoding a polypeptide with biological activity.
PT  useful for treating a disease or disorder, e.g. osteoarthritis, burns,
PT  CNS and peripheral disease, stroke, autoimmune disorders, viral
PT  infection, or cancer.
XX
XX  Claim 20; SEQ ID NO 937; 500pp; English.
PS
XX
XX  The invention describes a new isolated polynucleotide (I) encoding a
CC  polypeptide with biological activity comprising: a nucleotide sequence of
CC  SEQ ID NOS: 1-567 (fully defined); a nucleotide sequence that hybridizes
CC  to the sequence of (i) under stringent hybridization conditions; or a
CC  nucleotide sequence having greater than 99% sequence identity with the
CC  sequence of (i). Also described are: a(n) (expression)vector comprising
CC  (I); a host cell genetically engineered to comprise (I) operatively,
CC  associated with a regulatory sequence that modulates expression of the
CC  polynucleotide in the host cell; an isolated polypeptide comprising a
CC  sequence of SEQ ID NOS: 568-1134 (fully defined), where the polypeptide
CC  is: a polypeptide encoded by (I); or a polypeptide encoded by a
CC  polynucleotide hybridizing under stringent conditions with any one of SEQ
CC  ID NOS: 1-567; a composition comprising the polypeptide of (3) and a
CC  carrier; an antibody directed against the polypeptide of (3); a method
CC  for detecting (I) in a sample; a method for detecting the polypeptide of
CC  (3) in a sample; a method for identifying a compound that binds to the
CC  polypeptide of (3); a method of producing the polypeptide of (3); and a
CC  collection of polynucleotides, where the collection comprising of at
CC  least one of SEQ ID NOS: 1-567. (I) is a polynucleotide comprising any of
CC  the sequences of SEQ ID NOS: 1-567 encoding a polypeptide with biological
CC  activity, which comprises any of the amino acid sequence of SEQ ID NOS:
CC  568-1134. All sequences are fully defined in the specification. The
CC  sequences and methods are useful in diagnostics, forensic, and gene
CC  mapping, in identifying of mutations responsible for genetic disorders or
CC  other traits, in assessing biodiversity, and for producing many other
CC  types of data and products dependent on DNA and amino acid sequences. The
CC  composition and method are useful for treating a disease or disorder,
CC  e.g. osteoporosis, osteoarthritis, periodontal disease, burns, CNS and
CC  peripheral disease, Alzheimer's disease, Parkinson's disease, stroke,
CC  autoimmune disorders, viral infection, or cancer. This is the amino acid
CC  sequence of a novel polypeptide of the invention.
XX
XX  Sequence 679 AA;
SQ
XX
Query Match 75.8%; Score 50; DB 9; Length 679;
Best Local Similarity 83.3%; Pred. No. 6.6;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  4 SELNRVIOQLRS 15
DB  512 SELNRLIQIRRS 523
    |||||:|||||
    |||||:|||||

RESULT*14
AEA24282
ID  AEA24282 standard; peptide; 20 AA.
XX
XX  AEA24282;
AC
XX
XX  15-DEC-2005 (first entry)
DT
XX

```

DE Human protein-protein complex forming polypeptide fragment, SEQ ID 348.  
 XX  
 KW Cytostatic; Vasotropic; Cerebroprotective; Immunosuppressive;  
 KW Antidiabetic; Cardiant; Neuroprotective; Antiasthmatic; Antiinflammatory;  
 KW Antibacterial; Osteopathic; Anorectic; Virucide; Anti-HIV; Nephrotropic;  
 KW Antiartherosclerotic; Muscular-Gen.; Protein interaction;  
 KW Protein microarray; Cancer; Familial adenomatous polyposis;  
 KW Gastrointestinal-gen.; Ischemia; cerebrovascular ischemia;  
 KW autoimmune disease; diabetes; heart disease; neurodegenerative disease;  
 KW asthma; inflammation; sepsis; osteoporosis; obesity; Viral infection;  
 KW acquired immune deficiency syndrome; glomerulonephritis; atherosclerosis;  
 KW muscular dystrophy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2005222029-A1.  
 XX  
 PD 06-OCT-2005.  
 XX  
 PF 07-MAR-2005; 2005US-00075234.  
 XX  
 PR 04-JAN-2001; 2001US-0259571P.  
 PR 04-JAN-2001; 2001US-0259573P.  
 PR 14-MAR-2001; 2001US-0276259P.  
 PR 15-MAR-2001; 2001US-0276179P.  
 PR 19-MAR-2001; 2001US-0277013P.  
 PR 16-APR-2001; 2001US-0284095P.  
 PR 17-APR-2001; 2001US-0284220P.  
 PR 17-APR-2001; 2001US-0284404P.  
 PR 19-APR-2001; 2001US-0285324P.  
 PR 30-APR-2001; 2001US-0287513P.  
 PR 10-JUL-2001; 2001US-0304101P.  
 PR 23-JUL-2001; 2001US-0307233P.  
 PR 22-OCT-2001; 2001US-0347829P.  
 PR 25-OCT-2001; 2001US-034818P.  
 PR 04-JAN-2002; 2002US-00035344.  
 PR 07-JAN-2002; 2002US-0346384P.  
 PR 17-JAN-2002; 2002US-0349843P.  
 PR 06-FEB-2002; 2002US-0354899P.  
 PR 14-MAR-2002; 2002US-00098979.  
 PR 14-MAR-2002; 2002US-00099924.  
 PR 18-MAR-2002; 2002US-00100503.  
 PR 15-APR-2002; 2002US-00122573.  
 PR 17-APR-2002; 2002US-00124550.  
 PR 17-APR-2002; 2002US-00124767.  
 PR 18-APR-2002; 2002US-00125639.  
 PR 29-APR-2002; 2002US-00135802.  
 XX  
 PA (MYRI-) MYRIAD GENETICS INC.  
 XX  
 XX Bartel P, Cimbora D, Sugiyama J, Wettstein DA, Heichman K;  
 PI  
 XX WPI; 2005-664172/68.  
 DR  
 XX  
 XX New isolated protein complex having a first protein interacting with a  
 PT second protein, useful for treating or preventing, e.g. cancer, ischemia,  
 PT stroke, autoimmune diseases, diabetes, coronary heart disease, or asthma.  
 XX  
 PS Disclosure; SEQ ID NO 348; 198pp; English.  
 XX  
 CC The invention relates to a novel isolated protein complex having a first  
 CC protein interacting with a second protein. The invention further  
 CC comprises a protein microarray comprising the protein complex; a method  
 CC for selecting modulators of the protein complex; a method of selecting  
 CC modulators of an interaction between a first protein and a second protein  
 CC ; and the treating and/or preventing of diseases and disorders associated  
 CC with the protein complexes. The protein complexes are useful in screening  
 CC assays for identifying compounds effective in modulating the protein  
 CC complexes, and in treating and/or preventing diseases and disorders  
 CC associated with the protein complexes. The diseases and disorders include  
 CC cancer, adenomatous polyposis, ischemia, stroke, autoimmune diseases,  
 CC cancer, coronary heart disease, neurodegenerative diseases, asthma,  
 CC inflammatory disorders, sepsis, osteoporosis, obesity, viral infection,

CC AIDS, glomerulonephritis, atherosclerosis, and muscular dystrophy. This  
 CC sequence represents polypeptide fragment of a human protein which forms  
 CC part of a protein-protein complex of the invention.  
 XX  
 SQ Sequence 20 AA;  
 Query Match 71.2%; Score 47; DB 9; Length 20;  
 Best Local Similarity 75.0%; Pred. No. 0.47;  
 Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 SELNRVIQLRLS 15  
 : : : : :  
 Db 2 ABINRMQLRLS 13  
 RESULT 15  
 ADR98933  
 ID ADR98933 standard; protein; 92 AA.  
 XX  
 AC ADR98933;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Lung specific gene splice variant encoded protein #105.  
 XX  
 KW cytostatic; gene therapy; vaccine; lung; diagnosis; cancer;  
 KW non-cancerous lung disease; lung tissue; antagonist; gene therapy;  
 KW transgenic animal; splice variant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004074430-A2.  
 XX  
 PD 02-SEP-2004.  
 XX  
 PF 08-DEC-2003; 2003WO-US038896.  
 XX  
 PR 06-DEC-2002; 2002US-0431307P.  
 PR 06-DEC-2002; 2002US-0431510P.  
 PR 06-DEC-2002; 2002US-0431516P.  
 XX  
 PA (DIAD-) DIADEXUS INC.  
 XX  
 XX Macina RA, Turner LR, Sun Y, Liu S;  
 PI  
 XX WPI; 2004-635553/61.  
 DR  
 DR N-PSDB; ADR98793.  
 XX  
 XX New isolated human lung specific nucleic acid molecule, useful for  
 PT identifying, diagnosing, monitoring, staging, imaging and treating lung  
 PT cancer and non-cancerous diseases of the lung.  
 XX  
 PS Claim 1; SEQ ID NO 218; 542pp; English.  
 XX  
 CC The invention relates to a new isolated lung specific nucleic acid  
 CC molecule (1) comprising any of 113 fully defined nucleotide sequences  
 CC given in the specification, their encoded protein sequences, sequences  
 CC selectively hybridizing to the nucleotide sequences or a sequence having  
 CC at least 60% identity to the nucleotide sequences. The methods and  
 CC compositions of the present invention are useful for identifying,  
 CC diagnosing, monitoring, staging, imaging and treating lung cancer and non  
 CC -cancerous diseases of the lung. They are also used for identifying lung  
 CC tissue, monitoring and identifying and/or designing antagonists of the  
 CC polypeptide of the invention, gene therapy, production of transgenic  
 CC animals and production of engineered lung tissue for treatment and  
 CC research. Lung specific genes (LSGs) were identified by a systematic  
 CC analysis of gene expression data in the LiferSEQ Gold database using the  
 CC data mining software package candidate lead automatic search program  
 CC (CLASP). Genes were grouped into gene bins where each bin is a cluster of  
 CC sequences grouped together where they share a common contig.  
 CC Differentially expressed tissue-specific genes were selected based on the  
 CC percentage level in the targeted tissue versus all the other tissues. The  
 CC expression levels for each gene in libraries of normal tissues or non-



CC tumour tissues from cancer patients were compared with the expression  
 CC levels in tissue libraries associated with tumour or disease. This  
 CC sequence represents a protein of the invention.

XX

SQ Sequence 92 AA;

Query Match 71.2%; Score 47; DB 8; Length 92;  
 Best Local Similarity 75.0%; Pred. No. 2.5;  
 Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLS 15

Db 21 SENNRMIQLRLA 32

Search completed: July 5, 2006, 18:43:56  
 Job time : 140.029 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 117.126 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-9

Perfect score: 66

Sequence: 1 XXSELRVQLRLSI 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	59	89.4	16	5	US-10-774-928-9
2	55	83.3	643	5	US-10-480-848A-3
3	55	83.3	644	5	US-10-723-860-366
4	50	75.8	398	4	US-10-104-047-3388
5	50	75.8	398	6	US-11-072-512-3388
6	50	75.8	511	4	US-10-015-115-46
7	49	74.2	524	3	US-09-976-782-88
8	47	71.2	20	6	US-11-075-234-348
9	47	71.2	551	3	US-09-976-782-89
10	47	71.2	551	3	US-09-976-782-90
11	47	71.2	551	4	US-10-015-115-127
12	47	71.2	551	4	US-10-231-913-116
13	47	71.2	551	4	US-10-231-913-117
14	47	71.2	551	5	US-10-719-993-865
15	47	71.2	557	5	US-10-719-993-864
16	47	71.2	564	4	US-10-170-385-421
17	47	71.2	564	5	US-10-723-860-1619
18	47	71.2	564	5	US-10-756-149-5130
19	47	71.2	564	5	US-10-631-467-938
20	47	71.2	564	6	US-11-037-713-27
21	47	71.2	567	4	US-10-104-047-3772
22	47	71.2	567	6	US-11-072-512-3772
23	47	71.2	585	5	US-10-821-234-875
24	47	71.2	645	3	US-09-919-172-41
25	47	71.2	645	5	US-10-752-986-41
26	46	69.7	222	4	US-10-231-913-263
27	46	69.7	315	4	US-10-231-913-259

Sequence 261, App  
Sequence 1475, App  
Sequence 10, Appl  
Sequence 48, Appl  
Sequence 32, Appl  
Sequence 11, Appl  
Sequence 85, Appl  
Sequence 87, Appl  
Sequence 86, Appl  
Sequence 38, Appl  
Sequence 2446, App  
Sequence 264194,  
Sequence 124, App  
Sequence 85, Appl  
Sequence 115, App  
Sequence 123, App  
Sequence 83, Appl

28 46 69.7 318 4 US-10-231-913-261  
29 46 69.7 327 4 US-10-408-765A-1475  
30 46 69.7 521 3 US-09-976-782-10  
31 46 69.7 521 4 US-10-015-115-48  
32 46 69.7 526 4 US-10-231-913-32  
33 46 69.7 529 4 US-10-250-613-11  
34 46 69.7 534 3 US-09-976-782-86  
35 46 69.7 534 3 US-09-976-782-87  
36 46 69.7 534 4 US-10-231-913-86  
37 46 69.7 534 4 US-10-231-913-87  
38 46 69.7 534 5 US-10-367-057-38  
39 46 69.7 608 4 US-10-094-749-2446  
40 45 68.2 95 4 US-10-425-115-264194  
41 44 66.7 441 4 US-10-015-115-124  
42 44 66.7 441 4 US-10-231-913-85  
43 44 66.7 441 4 US-10-231-913-115  
44 44 66.7 523 4 US-10-015-115-123  
45 44 66.7 523 4 US-10-231-913-83

#### ALIGNMENTS

##### RESULT 1

US-10-774-928-9

; Sequence 9, Application US/10774928

; Publication No. US20040241168A1

; GENERAL INFORMATION:

; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission

; FILE REFERENCE: Psoriasis

; FILE REFERENCE: 302178

; CURRENT APPLICATION NUMBER: US/10/774,928

; CURRENT FILING DATE: 2004-02-09

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 9

; LENGTH: 16

; TYPE: PRT

; ORGANISM: Leishmania

; FEATURE:

; NAME/KEY: Xaa

; LOCATION: (1)..(3)

; OTHER INFORMATION: Xaa is an unknown amino acid.

US-10-774-928-9

Query Match 89.4%; Score 59; DB 5; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.0042;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVQLRLSI 16

Db 4 SELNRVQLRLSI 16

##### RESULT 2

US-10-480-848A-3

; Sequence 3, Application US/10480848A

; Publication No. US20040219597A1

; GENERAL INFORMATION:

; APPLICANT: B.R.A.H.M.S Aktiengesellschaft

; TITLE OF INVENTION: Use of soluble cytochrome c fragments in diagnosis and

; FILE REFERENCE: 3537PCTNAT

; CURRENT APPLICATION NUMBER: US/10/480,848A

; CURRENT FILING DATE: 2003-12-15

; PRIOR APPLICATION NUMBER: DE 101 30 985.6

; PRIOR FILING DATE: 2001-06-27

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 643

; TYPE: PRT

```
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; TITLE: L.Johnson et al., Structure of a gene for the human
; TITLE: epidermal 67-kDa keratin
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 1985
; ISSUE: 82
; PAGES: 1896-1900
US-10-480-848A-3

Query Match      83.3%; Score 55; DB 5; Length 643;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 SELNRVIOQLRS 15
Db      398 SELNRVIOQLRS 409
|||||:|||||

RESULT 3
US-10-723-860-366
; Sequence 366, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8193
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 366
; LENGTH: 644
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-366

Query Match      83.3%; Score 55; DB 5; Length 644;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 SELNRVIOQLRS 15
Db      399 SELNRVIOQLRS 410
|||||:|||||

RESULT 4
US-10-104-047-3388
; Sequence 3388, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3388
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3388

Query Match      75.8%; Score 50; DB 4; Length 398;
Best Local Similarity 83.3%; Pred. No. 4.6;
Matches 10; Indels 0; Gaps 0;
```

```
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      4 SELNRVIOQLRS 15
Db      231 SELNRVIOQLRS 242
|||||:|||||

RESULT 5
US-11-072-512-3388
; Sequence 3388, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3388
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3388

Query Match      75.8%; Score 50; DB 6; Length 398;
Best Local Similarity 83.3%; Pred. No. 4.6;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      4 SELNRVIOQLRS 15
Db      231 SELNRVIOQLRS 242
|||||:|||||

RESULT 6
US-10-015-115-46
; Sequence 46, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xisaojia
; APPLICANT: Kekuda, Ramesha
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shimkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
```

US-10-015-115-46

TITLE OF INVENTION: Using the Same  
FILE REFERENCE: 21402-211  
CURRENT APPLICATION NUMBER: US/10/015,115  
CURRENT FILING DATE: 2002-09-23  
PRIOR APPLICATION NUMBER: 60/248,153  
PRIOR FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: 60/249,598  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/264,240  
PRIOR FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/266,127  
PRIOR FILING DATE: 2001-02-02  
PRIOR APPLICATION NUMBER: 60/269,562  
PRIOR FILING DATE: 2001-02-16  
PRIOR APPLICATION NUMBER: 60/304,348  
PRIOR FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/309,261  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/313,283  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 205  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 46  
LENGTH: 511  
TYPE: PRT  
ORGANISM: Homo sapiens

Query Match 75.8%; Score 50; DB 4; Length 511;  
Best Local Similarity 83.3%; Pred. No. 6;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRS 15  
Db 344 SELNRIOQLRS 355

RESULT 7

US-09-976-782-88  
Sequence 88, Application US/09976782  
Publication No. US20030190715A1  
GENERAL INFORMATION:  
APPLICANT: Grosse et al  
TITLE OF INVENTION: No. US20030190715A1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-157  
CURRENT APPLICATION NUMBER: US/09/976,782  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: 60/240,113  
PRIOR FILING DATE: 2000-10-12  
PRIOR APPLICATION NUMBER: 60/240,662  
PRIOR FILING DATE: 2000-10-16  
PRIOR APPLICATION NUMBER: 60/240,732  
PRIOR FILING DATE: 2000-10-16  
PRIOR APPLICATION NUMBER: 60/240,625  
PRIOR FILING DATE: 2000-10-16  
PRIOR APPLICATION NUMBER: 60/240,703  
PRIOR FILING DATE: 2000-10-16  
PRIOR APPLICATION NUMBER: 60/241,190  
PRIOR FILING DATE: 2000-10-16  
PRIOR APPLICATION NUMBER: 60/240,637  
PRIOR FILING DATE: 2000-10-16  
PRIOR APPLICATION NUMBER: 60/240,669  
PRIOR FILING DATE: 2000-10-16  
PRIOR APPLICATION NUMBER: 60/262,455  
PRIOR FILING DATE: 2001-01-18  
PRIOR APPLICATION NUMBER: 60/240,648  
PRIOR FILING DATE: 2000-10-16  
NUMBER OF SEQ ID NOS: 127  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 88  
LENGTH: 524  
TYPE: PRT  
ORGANISM: Homo sapiens

US-09-976-782-88

Query Match 74.2%; Score 49; DB 3; Length 524;  
Best Local Similarity 83.3%; Pred. No. 9.1;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRS 15  
Db 365 SELNRVIOQLRS 376

RESULT 8

US-11-075-234-348  
Sequence 348, Application US/11075234  
Publication No. US20050222029A1  
GENERAL INFORMATION:  
APPLICANT: Myriad Genetics  
APPLICANT: Bartel, Paul  
APPLICANT: Cimborra, Daniel  
APPLICANT: Sugiyama, Janice  
APPLICANT: Wettstein, Daniel A  
APPLICANT: Heichman, Karen  
TITLE OF INVENTION: Compositions and Methods for Treating Diseases  
FILE REFERENCE: 1835.01  
CURRENT APPLICATION NUMBER: US/11/075,234  
CURRENT FILING DATE: 2005-03-08  
PRIOR APPLICATION NUMBER: US 10/135802  
PRIOR FILING DATE: 2002-04-29  
PRIOR APPLICATION NUMBER: US 10/035344  
PRIOR FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: US 10/098979  
PRIOR FILING DATE: 2002-03-13  
PRIOR APPLICATION NUMBER: US 10/099924  
PRIOR FILING DATE: 2002-04-14  
PRIOR APPLICATION NUMBER: US 10/122573  
PRIOR FILING DATE: 2002-04-15  
PRIOR APPLICATION NUMBER: US 10/124550  
PRIOR FILING DATE: 2002-04-17  
PRIOR APPLICATION NUMBER: US 10/124767  
PRIOR FILING DATE: 2002-04-17  
PRIOR APPLICATION NUMBER: US 10/125639  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 10/100503  
PRIOR FILING DATE: 2002-04-18  
NUMBER OF SEQ ID NOS: 469  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 348  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Homo sapiens

Query Match 71.2%; Score 47; DB 6; Length 20;  
Best Local Similarity 75.0%; Pred. No. 0.58;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRS 15  
Db 2 AELNRVIOQLRS 13

RESULT 9

US-09-976-782-89  
Sequence 89, Application US/09976782  
Publication No. US20030190715A1  
GENERAL INFORMATION:  
APPLICANT: Grosse et al  
TITLE OF INVENTION: No. US20030190715A1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-157  
CURRENT APPLICATION NUMBER: US/09/976,782  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: 60/240,113  
PRIOR FILING DATE: 2000-10-12

; PRIOR APPLICATION NUMBER: 60/240,662  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,732  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,625  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,703  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/241,190  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,637  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,669  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/262,455  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/240,648  
; PRIOR FILING DATE: 2000-10-16  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 89  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-976-782-89

Query Match 71.2%; Score 47; DB 3; Length 551;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIQRLRS 15  
||:|||||:  
Db 368 SEMNRMIQRLRA 379

RESULT 10  
US-09-976-782-90  
; Sequence 90, Application US/09976782  
; Publication No. US20030190715A1  
; GENERAL INFORMATION:  
; APPLICANT: Grosse et al  
; TITLE OF INVENTION: No. US20030190715A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-157  
; CURRENT APPLICATION NUMBER: US/09/976,782  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,113  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: 60/240,662  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,732  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,625  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,703  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/241,190  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,637  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/240,669  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR APPLICATION NUMBER: 60/262,455  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/240,648  
; PRIOR FILING DATE: 2000-10-16  
; NUMBER OF SEQ ID NOS: 127  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 90  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-782-90

Query Match 71.2%; Score 47; DB 3; Length 551;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIQRLRS 15  
||:|||||:  
Db 368 SEMNRMIQRLRA 379

RESULT 11  
US-10-015-115-127  
; Sequence 127, Application US/10015115  
; Publication No. US20030207800A1  
; GENERAL INFORMATION:  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Shenoy, Suresh G  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Kekuda, Ramesha  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Shimkete, Richard A  
; APPLICANT: Taupier, Raymond J  
; APPLICANT: Li, Li  
; APPLICANT: Padigaru, Muralidhara  
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
; FILE REFERENCE: 21402-211  
; CURRENT APPLICATION NUMBER: US/10/015,115  
; CURRENT FILING DATE: 2002-09-23  
; PRIOR APPLICATION NUMBER: 60/248,153  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: 60/249,598  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/264,240  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/266,127  
; PRIOR FILING DATE: 2001-02-02  
; PRIOR APPLICATION NUMBER: 60/269,562  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: 60/304,348  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/309,261  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/313,283  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 205  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 127  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-015-115-127

Query Match 71.2%; Score 47; DB 4; Length 551;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIQRLRS 15  
||:|||||:  
Db 368 SEMNRMIQRLRA 379

RESULT 12  
US-10-231-913-116  
; Sequence 116, Application US/10231913  
; Publication No. US20040005576A1  
; GENERAL INFORMATION:  
; APPLICANT: Guo, Xiaojia S.  
; APPLICANT: Li, Li  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkete, Richard A.

; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Edinger, Schlomit  
; APPLICANT: Peyman, John A.  
; APPLICANT: Stone, David J.  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Boldog, Ference L.  
; APPLICANT: Colman, Steven D.  
; APPLICANT: Eisen, Andrew J.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Spaderna, Steven K.  
; APPLICANT: Zerhusen, Bryan D.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-216  
; CURRENT APPLICATION NUMBER: US/10/231.913  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: 60/251,660  
; PRIOR FILING DATE: 2000-12-06  
; PRIOR APPLICATION NUMBER: 60/255,029  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: 60/260,326  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: 60/263,800  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/269,942  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/286,183  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/313,627  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/318,712  
; PRIOR FILING DATE: 2001-09-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 116  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-231-913-116

Query Match 71.2%; Score 47; DB 4; Length 551;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRLS 15  
Db 368 SEMNRMIQLRLA 379  
||:|||||:

RESULT 13  
US-10-231-913-117  
; Sequence 117, Application US/10231913  
; Publication No. US20040005576A1  
; GENERAL INFORMATION:  
; APPLICANT: Guo, Xiaojia S.  
; APPLICANT: Li, Li  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Edinger, Schlomit

; APPLICANT: Peyman, John A.  
; APPLICANT: Stone, David J.  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Boldog, Ference L.  
; APPLICANT: Colman, Steven D.  
; APPLICANT: Eisen, Andrew J.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Spaderna, Steven K.  
; APPLICANT: Zerhusen, Bryan D.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-216  
; CURRENT APPLICATION NUMBER: US/10/231.913  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: 60/251,660  
; PRIOR FILING DATE: 2000-12-06  
; PRIOR APPLICATION NUMBER: 60/255,029  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: 60/260,326  
; PRIOR FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: 60/263,800  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/269,942  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/286,183  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/313,627  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/318,712  
; PRIOR FILING DATE: 2001-09-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 117  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-231-913-117

Query Match 71.2%; Score 47; DB 4; Length 551;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRLS 15  
Db 368 SEMNRMIQLRLA 379  
||:|||||:

RESULT 14  
US-10-719-993-865  
; Sequence 865, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 865  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-993-865

Query Match 71.2%; Score 47; DB 5; Length 551;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SELNRVIOQLRLS 15  
||:|||||:

Db 368 SEMNRMIQRLRA 379

## RESULT 15

US-10-719-993-864  
; Sequence 864, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 864  
; LENGTH: 557  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-993-864

Query Match 71.2%; Score 47; DB 5; Length 557;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRS 15

||:|||:

Db 374 SEMNRMIQRLRA 385

Search completed: July 5, 2006, 20:52:01  
Job time : 118.126 secs





; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 55  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-11-178-538-55

Query Match 56.1%; Score 37; DB 7; Length 309;  
Best Local Similarity 58.3%; Pred. No. 30;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 YSELNRVQLRLR 14  
| | | | | : | | :  
Db 68 YEELERVVLRK 79

## RESULT 3

US-11-178-538-62  
; Sequence 62, Application US/11178538  
; Publication No. US20060110792A1  
; GENERAL INFORMATION:  
; APPLICANT: Pauech, Mark H  
; APPLICANT: Price, Laura A  
; TITLE OF INVENTION: POTASSIUM CHANNELS, NUCLEOTIDE SEQUENCES ENCODING THEM,  
; TITLE OF INVENTION: AND METHODS OF USING SAME  
; FILE REFERENCE: 01142.0122 SEQUENCE LISTING  
; CURRENT APPLICATION NUMBER: US/11/178,538  
; CURRENT FILING DATE: 2005-07-12  
; PRIOR APPLICATION NUMBER: US/08/816,011  
; PRIOR FILING DATE: 1997-03-11  
; PRIOR APPLICATION NUMBER: 07/332,312  
; PRIOR FILING DATE: 1994-10-31  
; PRIOR APPLICATION NUMBER: PCT/US95/14364  
; PRIOR FILING DATE: 1995-10-25  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 62  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-11-178-538-62

Query Match 56.1%; Score 37; DB 7; Length 309;  
Best Local Similarity 58.3%; Pred. No. 30;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 YSELNRVQLRLR 14  
| | | | | : | | :  
Db 68 YEELERVVLRK 79

## RESULT 4

US-10-449-902-40943  
; Sequence 40943, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 40943  
; LENGTH: 533  
; TYPE: PRT

; ORGANISM: Oryza sativa  
US-10-449-902-40943

Query Match 56.1%; Score 37; DB 6; Length 533;  
Best Local Similarity 81.8%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ELNRVQLRLRS 15  
| | | | | : | | :  
Db 308 ELLRVQLRLNS 318

## RESULT 5

US-10-953-349-24947  
; Sequence 24947, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 24947  
; LENGTH: 141  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-953-349-24947

Query Match 54.5%; Score 36; DB 6; Length 141;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 YSELNRVQLRLRSI 16  
| | | | | : | | :  
Db 103 YGEMNRLSILQOI 116

## RESULT 6

US-10-953-349-24946  
; Sequence 24946, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 24946  
; LENGTH: 147  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-953-349-24946

Query Match 54.5%; Score 36; DB 6; Length 147;  
Best Local Similarity 50.0%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 YSELNRVQLRLRSI 16  
| | | | | : | | :  
Db 109 YGEMNRLSILQOI 122

## RESULT 7

US-10-953-349-24945  
; Sequence 24945, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:

```
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 24945
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Glycine max
; US-10-953-349-24945

Query Match 54.5%; Score 36; DB 6; Length 149;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YSELNRVQIQLRSI 16
Db 111 YGENRLISILQOI 124

RESULT 8
US-10-953-349-37808
; Sequence 37808, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 37808
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; US-10-953-349-37808

Query Match 54.5%; Score 36; DB 6; Length 163;
Best Local Similarity 46.2%; Pred. No. 23;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 3 YSELNRVQIQLRS 15
Db 129 FYEYNRVKVRVT 141

RESULT 9
US-10-449-902-44210
; Sequence 44210, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOR-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 44210
; LENGTH: 663
; TYPE: PRT
```

```
; ORGANISM: Oryza sativa
; US-10-449-902-44210

Query Match 54.5%; Score 36; DB 6; Length 663;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 YSELNRVQIQLRSI 16
Db 186 YSEDHRVVOQMTDI 199

RESULT 10
US-10-953-349-26148
; Sequence 26148, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26148
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Triticum aestivum
; US-10-953-349-26148

Query Match 53.0%; Score 35; DB 6; Length 398;
Best Local Similarity 70.0%; Pred. No. 88;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ELNRVQIQLR 14
Db 367 ELKRVQIQRVK 376

RESULT 11
US-10-953-349-26147
; Sequence 26147, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26147
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Triticum aestivum
; US-10-953-349-26147

Query Match 53.0%; Score 35; DB 6; Length 402;
Best Local Similarity 70.0%; Pred. No. 88;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ELNRVQIQLR 14
Db 371 ELKRVQIQRVK 380

RESULT 12
US-10-953-349-26146
; Sequence 26146, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
```

```

; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31041
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-31041

Query Match      51.5%; Score 34; DB 6; Length 216;
Best Local Similarity 70.0%; Pred. No. 68;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 YSELNRVRIQR 12
      ||| |||:
Db      179 YSSLERVIIQ 188

RESULT 15
US-10-449-902-32497
; Sequence 32497, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32497
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-32497

Query Match      51.5%; Score 34; DB 6; Length 235;
Best Local Similarity 41.7%; Pred. No. 74;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      3 YSELNRVRIQRLR 14
      |:::|::|:
Db      17 YAKIDRVLTFR 28

Search completed: July 5, 2006, 20:53:04
Job time : 9.16505 secs

```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 36.0388 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-9

Perfect score: 66

Sequence: 1 XXYSLNRVQLRLSI 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/iaa/5 COMB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/iaa/6 COMB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/iaa/7 COMB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/iaa/H COMB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/iaa/PCTUS COMB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/iaa/RE COMB.pep.\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/iaa/backfilesl.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	89.4	16	2	US-09-809-003A-9
2	55	83.3	637	2	US-09-949-016-8152
3	55	83.3	643	2	US-09-538-092-844
4	50	75.8	398	2	US-10-104-047-3380
5	47	71.2	551	1	US-09-067-351-2
6	47	71.2	551	2	US-09-360-490-2
7	47	71.2	564	2	US-09-949-016-6628
8	47	71.2	567	2	US-10-104-047-3772
9	47	71.2	569	2	US-09-949-016-11035
10	47	71.2	569	2	US-09-949-016-11036
11	47	71.2	645	2	US-09-919-172-41
12	46	69.7	608	2	US-10-094-749-2446
13	42	63.6	469	2	US-09-077-606-3
14	42	63.6	476	2	US-09-949-016-9096
15	42	63.6	546	1	US-09-067-351-1
16	42	63.6	546	2	US-09-360-490-1
17	41	62.1	149	2	US-09-270-767-58244
18	41	62.1	313	2	US-09-270-767-42921
19	40	60.6	235	2	US-09-620-405B-487
20	40	60.6	235	2	US-09-604-287A-487
21	40	60.6	235	2	US-09-834-759-487
22	40	60.6	235	2	US-10-076-622-487
23	40	60.6	235	2	US-10-124-805-487
24	40	60.6	274	2	US-09-949-016-9095
25	40	60.6	366	2	US-09-252-991A-16760
26	40	60.6	384	2	US-09-134-000C-5839

```

27 40 60.6 493 2 US-09-949-016-6203 Sequence 6203, Ap
28 40 60.6 505 2 US-09-620-405B-478 Sequence 478, App
29 40 60.6 505 2 US-09-620-405B-485 Sequence 485, App
30 40 60.6 505 2 US-09-604-287A-478 Sequence 478, App
31 40 60.6 505 2 US-09-604-287A-485 Sequence 485, App
32 40 60.6 505 2 US-09-834-759-478 Sequence 478, App
33 40 60.6 505 2 US-09-834-759-485 Sequence 485, App
34 40 60.6 505 2 US-09-590-751A-478 Sequence 478, App
35 40 60.6 505 2 US-09-551-621-478 Sequence 478, App
36 40 60.6 505 2 US-09-551-621A-478 Sequence 478, App
37 40 60.6 505 2 US-10-076-622-478 Sequence 485, App
38 40 60.6 505 2 US-10-076-622-485 Sequence 485, App
39 40 60.6 505 2 US-10-124-805-478 Sequence 478, App
40 40 60.6 505 2 US-10-124-805-485 Sequence 485, App
41 40 60.6 528 2 US-09-949-016-9678 Sequence 9678, A
42 39 59.1 368 2 US-09-252-991A-32075 Sequence 32075, A
43 39 59.1 952 2 US-09-248-796A-20878 Sequence 20878, A
44 37 56.1 54 2 US-09-513-999C-5139 Sequence 5139, Ap
45 37 56.1 318 2 US-09-107-532A-5738 Sequence 5738, Ap

```

#### ALIGNMENTS

```

RESULT 1
US-09-809-003A-9
; Sequence 9, Application US/09809003A
; Patent No. 6673351
; GENERAL INFORMATION:
; APPLICANT: O'Daly, Jose
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission
; TITLE OF INVENTION: Psoriasis
; FILE REFERENCE: 302178
; CURRENT APPLICATION NUMBER: US/09/809,003A
; CURRENT FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 9
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Leishmania
; FEATURE:
; NAME/KEY: Xaa
; LOCATION: (1)..(3)
; OTHER INFORMATION: Xaa is an unknown amino acid.
US-09-809-003A-9

```

```

Query Match 89.4%; Score 59; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00084;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 SELNRVQLRLSI 16
|||||
Db 4 SELNRVQLRLSI 16

```

```

RESULT 2
US-09-949-016-8152
; Sequence 8152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8152
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8152

Query Match      83.3%; Score 55; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 SELNRVQLRLS 15
Db      399 SELNRVQLRLS 410
|||||:|||||

RESULT 3
US-09-538-092-844
; Sequence 844, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatchFormatter Version 0.9
; SEQ ID NO 844
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P04264
US-09-538-092-844

Query Match      83.3%; Score 55; DB 2; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 SELNRVQLRLS 15
Db      398 SELNRVQLRLS 409
|||||:|||||

RESULT 4
US-10-104-047-3388
; Sequence 3388, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3388
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3388

Query Match      83.3%; Score 50; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 0.95;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      4 SELNRVQLRLS 15
Db      231 SELNRVQLRLS 242
|||||:|||||

RESULT 5
US-09-067-351-2
; Sequence 2, Application US/09067351
; Patent No. 5994081
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,351
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
; CLONE: 2029060
US-09-067-351-2

Query Match      71.2%; Score 47; DB 1; Length 551;
Best Local Similarity 75.0%; Pred. No. 4.4;
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      4 SELNRVQLRLS 15
Db      368 SEMNRMIQLRA 379
|||||:|||||

RESULT 6
US-09-360-490-2
; Sequence 2, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
```

; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/360,490  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/067,351  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: amino acid  
; LENGTH: 551 amino acids  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: KERANOT02  
; CLONE: 2029060  
; US-09-360-490-2

Query Match 71.2%; Score 47; DB 2; Length 551;  
Best Local Similarity 75.0%; Pred. No. 4.4;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVRIQLRLS 15  
DB 368 SEMNRMIQLRLA 379

RESULT 7  
US-09-949-016-6628  
; Sequence 6628, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6628  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-949-016-6628

Query Match 71.2%; Score 47; DB 2; Length 564;

Best Local Similarity 75.0%; Pred. No. 4.5;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 4 SELNRVRIQLRLS 15  
DB 382 AEINRMQLRLS 393

RESULT 8  
US-10-104-047-3772  
; Sequence 3772, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3772  
; LENGTH: 567  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-104-047-3772

Query Match 71.2%; Score 47; DB 2; Length 567;  
Best Local Similarity 75.0%; Pred. No. 4.6;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVRIQLRLS 15  
DB 364 SEMNRMIQLRLA 375

RESULT 9  
US-09-949-016-11035  
; Sequence 11035, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11035  
; LENGTH: 569  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-949-016-11035

Query Match 71.2%; Score 47; DB 2; Length 569;  
Best Local Similarity 75.0%; Pred. No. 4.6;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVRIQLRLS 15  
DB 387 AEINRMQLRLS 398

RESULT 10  
US-09-949-016-11036

```

; Sequence 11036, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11036
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-11036

Query Match 71.2%; Score 47; DB 2; Length 569;
Best Local Similarity 75.0%; Pred. No. 4.6;
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLS 15
DB 387 AEINRMIQLRLS 398
      :|:|:|:|:|:|

RESULT 11
US-09-919-172-41
; Sequence 41, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673545 1798379CD1
US-09-919-172-41

Query Match 71.2%; Score 47; DB 2; Length 645;
Best Local Similarity 90.9%; Pred. No. 5.2;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 SELNRVQLRLR 14
DB 403 SELNRVQLRLQ 413
      |||:|:|:|:|:|:|

RESULT 12
US-10-094-749-2446
; Sequence 2446, Application US/10094749
; Patent No. 6979557
; GENERAL INFORMATION:
; APPLICANT: ISOGAL, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI

```



RESULT 14  
US-09-949-016-9096  
; Sequence 9096, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9096  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-9096

Query Match 63.6%; Score 42; DB 2; Length 476;  
Best Local Similarity 66.7%; Pred. No. 28;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRS 15  
DB 316 SEMNRAIQRLQA 327

RESULT 15  
US-09-067-351-1  
; Sequence 1, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,351  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 amino acids  
; TYPE: amino acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PANTUT02  
; CLONE: 1467090  
US-09-067-351-1

Query Match 63.6%; Score 42; DB 1; Length 546;  
Best Local Similarity 66.7%; Pred. No. 32;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 SELNRVIOQLRS 15  
DB 386 SEMNRAIQRLQA 397

Search completed: July 5, 2006, 19:24:27  
Job time : 36.0388 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 9.70874 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928a-10

Perfect score: 43

Sequence: 1 AQYEDYQA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	81.4	629	2	A29666	keratin, 65K type
2	35	81.4	629	2	S42629	keratin K3 - rabbit
3	35	81.4	636	2	S09246	coagulase precursor
4	35	81.4	729	2	S76065	hypothetical prote
5	35	81.4	1979	2	JW0059	mtpr protein - mo
6	34	79.1	730	2	T49736	hypothetical prote
7	32	74.4	163	2	G87551	hypothetical prote
8	32	74.4	305	2	E69946	hypothetical prote
9	32	74.4	522	2	S52216	vital proteinase -
10	32	74.4	603	2	H90061	hypothetical prote
11	32	74.4	1084	2	G71329	hypothetical prote
12	32	74.4	5255	2	T31677	bacitracin synthet
13	31	72.1	122	2	S24287	ig heavy chain V r
14	31	72.1	140	2	T30210	hypothetical prote
15	31	72.1	156	2	T32236	hypothetical prote
16	31	72.1	205	1	F71193	hypothetical prote
17	31	72.1	296	2	A70973	hypothetical prote
18	31	72.1	307	2	D69521	hypothetical prote
19	31	72.1	380	2	A82589	hypothetical prote
20	31	72.1	384	2	I61769	keratin 6d, type I
21	31	72.1	446	2	AH1957	hypothetical prote
22	31	72.1	503	2	T24985	hypothetical prote
23	31	72.1	506	2	AF2192	bifunctional purin
24	31	72.1	522	2	S67704	hypothetical prote
25	31	72.1	524	2	A23518	keratin, 57K type
26	31	72.1	534	2	I37942	keratin 4, type II
27	31	72.1	544	2	F81059	YbX/YbJW/VjP/VjD
28	31	72.1	564	1	KRHUEA	keratin 6a, type I
29	31	72.1	564	1	KRHUEB	keratin 6b, type I

30 31 72.1 564 2 D83211 probable malic enz  
31 31 72.1 564 2 I61770 keratin 6e, type I  
32 31 72.1 564 2 I61768 keratin 6c, type I  
33 31 72.1 564 2 I61771 keratin 6f, type I  
34 31 72.1 565 2 AC0184 malate dehydrogena  
35 31 72.1 574 2 B64901 malate dehydrogena  
36 31 72.1 574 2 F85728 NAD-linked malate  
37 31 72.1 574 2 C90889 NAD-linked malate  
38 31 72.1 577 2 A10672 NAD-linked malic e  
39 31 72.1 588 2 A82232 malate oxidoreduct  
40 31 72.1 645 2 A44861 keratin, 67K type  
41 31 72.1 755 2 B41836 amine oxidase (fla  
42 31 72.1 757 2 E84889 amine oxidase (cop  
43 31 72.1 1027 2 T27970 hypothetical prote  
44 31 72.1 1286 2 T18734 hypothetical prote  
45 31 72.1 1307 2 T25563 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

A29666

keratin, 65K type II cytoskeletal - human

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 20-Apr-2001

C:Accession: A29666

R:Klinge, E.M.; Sylvestre, Y.R.; Freedberg, I.M.; Blumenberg, M.

J. Mol. Evol. 24, 319-329, 1987

A>Title: Evolution of keratin genes: different protein domains evolve by different pathwa

A:Reference number: A29666; MUID:87254239; PMID:2439698

A:Accession: A29666

A:Molecule type: DNA

A:Residues: 1-629 <KLI>

A:Cross-references: UNIPARC:UPI0000161C35; GB:X05418; NID:934040; PIDN:CAA28991.1; PID:93

92; PIDN:CAA28994.1; PID:91335193; PIDN:CAA28995.1; PID:91335194; GB:X05421; NID:9340407;

A>Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 2;

A>Note: the complete translation is not annotated in GenBank entries HSKER65A, HSKER65B,

s mistranslated as a Met initiator codon

C:Genetics:

A:introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 81.4%; Score 35; DB 2; Length 629;

Best Local Similarity 87.5%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQA 8

DB 376 AQYEDYQA 383

##### RESULT 2

S42629

keratin K3 - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004

C:Accession: S42629

R:Wu, R.L.; Galvin, S.; Wu, S.K.; Xu, C.; Blumenberg, M.; Sun, T.T.

J. Cell Sci. 105, 303-316, 1993

A>Title: A 300 bp 5'-upstream sequence of a differentiation-dependent rabbit K3 keratin s

A:Reference number: S42629; MUID:94013038; PMID:7691837

A:Accession: S42629

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-629 <WUB>

A:Cross-references: UNIPROT:Q29426; UNIPARC:UPI0000087A39; EMBL:X74371; NID:9433957; PID:

C:Genetics:

A:introns: 200/3; 274/2; 294/3; 326/3; 381/3; 423/3; 497/2; 509/1

C:Superfamily: cytoskeletal keratin

Query Match 81.4%; Score 35; DB 2; Length 629;

```

Best Local Similarity 87.5%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQYEDYQAQ 8
    |||||
Db 361 AQYEDYQAQ 368

RESULT 3
S09246
coagulase precursor - Staphylococcus aureus (strain 8325-4)
C:Species: Staphylococcus aureus
A:Variety: strain 8325-4
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 09-Jul-2004
C:Accession: S09246
R:Phonindaeng, P.; O'Reilly, M.; Nowlan, P.; Bramley, A.J.; Foster, T.J.
Mol. Microbiol. 4, 393-404, 1990
A:Title: The coagulase of Staphylococcus aureus 8325-4. Sequence analysis and virulence
A:Reference number: S09246; MUID:90286916; PMID:2355852
A:Accession: S09246
A:Molecule type: DNA
A:Residues: 1-636 <PHO>
A:Cross-references: UNIPROT:Q53655; UNIPARC:UPI00000BC8C5; EMBL:X17679; NID:g46539; PIDN
C:Genetics:
A:Gene: coa
C:Keywords: extracellular protein
F:1-26/Domain: signal sequence #status predicted <SIG>
F:127-636/Product: coagulase #status predicted <MAT>

Query Match 81.4%; Score 35; DB 2; Length 636;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQYEDYQAQ 8
    :|||||
Db 105 SQYEDYKQ 112

RESULT 4
S76065
hypothetical protein slr0337 - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S76065
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76065
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-729 <KAN>
A:Cross-references: UNIPROT:Q55545; UNIPARC:UPI00000C105B; EMBL:D63999; GB:AB001339; NID
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 81.4%; Score 35; DB 2; Length 729;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQYEDYQAQ 8
    |||||
Db 69 AQLEDYQAQ 76

RESULT 5
JW0059
mtpr protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 09-Jul-2004
C:Accession: JW0059

```

```

R:Tsukahara, F.; Urakawa, I.; Hattori, M.; Hirai, M.; Ohba, K.; Yoshioka, T.; Sakaki, Y.;
J. Biochem. 123, 1055-1063, 1998
A:Title: Molecular characterization of the mouse mtpr gene, a homologue of human TPRD: t
A:Reference number: JW0059; MUID:98269034; PMID:9603993
A:Accession: JW0059
A:Molecule type: protein
A:Residues: 1-1979 <TSU>
A:Cross-references: UNIPROT:O88196; UNIPARC:UPI000001790C6
C:Comment: This protein has three units of tetratics-peptide repeats.
C:Genetics:
A:Gene: mtpr
A:Map position: 16C3.3-4
C:Superfamily: mtpr protein; tetratics-peptide repeat homology
F:1927-1976/Domain: RING finger homology <RRN>

Query Match 81.4%; Score 35; DB 2; Length 1979;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QYEDYQAQ 8
    |||||
Db 51 QYKDYQAQ 57

RESULT 6
T49736
hypothetical protein B24B19.80 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 15-Mar-2004
C:Accession: T49736
R:Schulze, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: 225022
A:Accession: T49736
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-730 <SCH>
A:Cross-references: UNIPARC:UPI000017B482; EMBL:AL356192; GSPDB:GN00116; NCSP:B24B19.80
A:Experimental source: BAC clone B24B19; strain OR74A
C:Genetics:
A:Gene: NCSP:B24B19.80
A:Map position: 6
A:Introns: 86/1
F:70-108/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 79.1%; Score 34; DB 2; Length 730;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 YEDYQAQ 8
    |||||
Db 597 YEDYQAQ 602

RESULT 7
G87551
hypothetical protein CC2440 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: G87551
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.I.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173696; PMID:11259647
A:Accession: G87551
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-163 <STO>
A:Cross-references: UNIPROT:Q9A5K7; UNIPARC:UPI00000C7721; GB:AE005673; NID:g13423985; P
C:Genetics:
A:Gene: CC2440

```

Query Match Best Local Similarity 74.4%; Score 32; DB 2; Length 163; Matches 6; Conservative 1; Mismatches 0; Gaps 0;	Db 309 RYDDYQAQ 315
RESULT 8 E69946 hypothetical protein yqbb - Bacillus subtilis C:Species: Bacillus subtilis C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004 C:Accession: E69946 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Eyring, J.; Fabbri, C.; Ferrari, E. Nature 390, 249-256, 1997 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F. Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toignoni, A.; Tosato, V.; Uchiyama T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis. A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. A:Reference number: A69580; MUID:98044033; PMID:9384377 A:Accession: E69946 A>Status: preliminary; nucleic acid sequence not shown; translation not shown A:Molecule type: DNA A:Residues: 1-305 <KUN> A:Cross-references: UNIPROT:P45918; UNIPARC:UPI000006079D; GB:Z99117; GB:AL009126; NID:9 A:Experimental source: strain 168 C:Genetics: A:Gene: yqbb	hypothetical protein SA2354 [imported] - Staphylococcus aureus (strain N315) C:Species: Staphylococcus aureus C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004 C:Accession: H90061 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; I C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus. A:Reference number: A89758; MUID:21311952; PMID:11418146 A:Accession: H90061 A>Status: preliminary A:Molecule type: DNA A:Residues: 1-603 <KUR> A:Cross-references: UNIPROT:Q99R71; UNIPARC:UPI00000548D5; GB:BA000018; PID:gl3702517; PJ A:Experimental source: strain N315 C:Genetics: A:Gene: SA2354
Query Match Best Local Similarity 74.4%; Score 32; DB 2; Length 603; Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;	Query Match 74.4%; Score 32; DB 2; Length 603; Best Local Similarity 62.5%; Pred. No. 1.4e+02; Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQVEDYQAQ 8 : : : : : Db 488 SQKDYAK 495	QY 1 AQVEDYQAQ 8 : : : : : Db 488 SQKDYAK 495
RESULT 11 G71329 hypothetical protein TP0408 - syphilis spirochete C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete) C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004 C:Accession: G71329 R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo rson, J.; Weidman, J.; Smith, H.O.; Venter, J.C. Science 281, 375-388, 1998 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete. A:Reference number: A71250; MUID:98332770; PMID:9665876 A:Accession: G71329 A>Status: preliminary; nucleic acid sequence not shown; translation not shown A:Molecule type: DNA A:Residues: 1-1084 <COL> A:Cross-references: UNIPROT:O83423; UNIPARC:UPI00000C0A6F; GB:AE001218; GB:AE000520; NID A:Experimental source: strain Nichols C:Genetics: A:Gene: TP0408	hypothetical protein TP0408 - syphilis spirochete C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete) C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004 C:Accession: G71329 R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo rson, J.; Weidman, J.; Smith, H.O.; Venter, J.C. Science 281, 375-388, 1998 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete. A:Reference number: A71250; MUID:98332770; PMID:9665876 A:Accession: G71329 A>Status: preliminary; nucleic acid sequence not shown; translation not shown A:Molecule type: DNA A:Residues: 1-1084 <COL> A:Cross-references: UNIPROT:O83423; UNIPARC:UPI00000C0A6F; GB:AE001218; GB:AE000520; NID A:Experimental source: strain Nichols C:Genetics: A:Gene: TP0408
Query Match Best Local Similarity 74.4%; Score 32; DB 2; Length 1084; Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;	Query Match 74.4%; Score 32; DB 2; Length 1084; Best Local Similarity 71.4%; Pred. No. 2.6e+02; Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 QVEDYQAQ 8 : : : : : Db 671 QYEDFSQ 677	QY 2 QVEDYQAQ 8 : : : : : Db 671 QYEDFSQ 677
RESULT 12 T31677 bacitracin synthetase 1 - Bacillus licheniformis C:Species: Bacillus licheniformis C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004 C:Accession: T31677 R:Kona, D.; Klens, A.; Schorgendorfer, K.; Marahiel, M.A. Chem. Biol. 4, 927-937, 1997 A:Title: The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecu A:Reference number: Z21058; MUID:98089193; PMID:9427658	RESULT 12 T31677 bacitracin synthetase 1 - Bacillus licheniformis C:Species: Bacillus licheniformis C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004 C:Accession: T31677 R:Kona, D.; Klens, A.; Schorgendorfer, K.; Marahiel, M.A. Chem. Biol. 4, 927-937, 1997 A:Title: The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecu A:Reference number: Z21058; MUID:98089193; PMID:9427658

A:Accession: T31677  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5255 <KON>  
A:Cross-references: UNIPROT:O68006; UNIPARC:UPI000013C49A; EMBL:AF007865; NID:G4464275;  
C:Genetics:  
A:Gene: bacA  
C:Keywords: carrier protein  
F:83-524/Domain: acetate-CoA ligase homology <ACU1>  
F:542-610/Domain: acyl carrier protein homology <ACP1>  
F:1134-1561/Domain: acetate-CoA ligase homology <ACU2>  
F:1583-1651/Domain: acyl carrier protein homology <ACP2>  
F:2171-2601/Domain: acetate-CoA ligase homology <ACU3>  
F:2619-2687/Domain: acyl carrier protein homology <ACP3>  
F:3195-3642/Domain: acetate-CoA ligase homology <ACU4>  
F:3662-3729/Domain: acyl carrier protein homology <ACP4>  
F:4712-5151/Domain: acetate-CoA ligase homology <ACU5>  
F:5169-5237/Domain: acyl carrier protein homology <ACP5>

Query Match 74.4%; Score 32; DB 2; Length 5255;  
Best Local Similarity 71.4%; Pred. No. 1.4e+03; Indels 0; Gaps 0;  
Matches 5; Conservative 2; Mismatches 0;

QY 2 QYEDYA 8  
|||:  
Db 2860 QYDYAE 2866

RESULT 13  
S24287  
Ig heavy chain V region (JS28/32) - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Jun-2000  
C:Accession: S24287  
R:Monchamont, B.  
submitted to the EMBL Data Library, September 1991  
A:Description: Cloning and sequencing of the cDNA coding for the variable regions of the  
A:Reference number: S24287  
A:Accession: S24287  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-122 <MON>  
A:Cross-references: UNIPARC:UPI0000116004; EMBL:X62706; NID:G51669; PIDN:CAA44588.1; PID  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 72.1%; Score 31; DB 2; Length 122;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDYA 7  
:|||:  
Db 102 SRYEDYA 108

RESULT 14  
T30210  
hypotheical protein 3 - Staphylococcus epidermidis  
C:Species: Staphylococcus epidermidis  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T30210  
R:Heilmann, C.; Huesain, M.; Peters, G.; Gotz, F.  
Mol. Microbiol. 24, 1013-1024, 1997.  
A:Title: Evidence for autolysin-mediated primary attachment of Staphylococcus epidermidis  
A:Reference number: 220779; PMID:97363715; PMID:9220008  
A:Accession: T30210  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-140 <HEI>  
A:Cross-references: UNIPROT:O33634; UNIPARC:UPI00000DC9D3; EMBL:U71377; NID:G2267238; PI  
C:Superfamily: hypotheical protein YJCF

Query Match 72.1%; Score 31; DB 2; Length 140;  
Best Local Similarity 83.3%; Pred. No. 49;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDYA 7  
|||:  
Db 36 QYEDYS 41

RESULT 15  
T32236  
hypotheical protein ZK488.10 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T32236  
R:Wilson, R.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid ZK488.  
A:Reference number: Z21138  
A:Accession: T32236  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-156 <WIL>  
A:Cross-references: UNIPROT:O17024; UNIPARC:UPI0000081A63; EMBL:AF022984; PIDN:AAB369957.1  
A:Experimental source: strain Bristol N2; clone ZK488  
C:Genetics:  
A:Gene: CESP:ZK488.10  
A:Map position: 5  
A:Introns: 74/3

Query Match 72.1%; Score 31; DB 2; Length 156;  
Best Local Similarity 62.5%; Pred. No. 55;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AQYEDYA 8  
:|||:  
Db 75 SQYSDYTQ 82

Search completed: July 5, 2006, 19:20:28  
Job time : 10.7087 secs

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 81.8641 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-10  
Perfect score: 43  
Sequence: 1 AQEDYQAQ 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	90.7	587	Q40275_KINRA	Q40275_kineococcus
2	38	86.4	948	Q35AN0_5BRAD	Q35AN0_bradyrhizob
3	37	86.0	212	Q8FRV4_COREF	Q8FRV4_corynebacte
4	36	83.7	198	Q9IB36_KATPE	Q9IB36_katsuwonus
5	35	81.4	210	Q3KDJ2_PSEPF	Q3KDJ2_pseudomonas
6	35	81.4	278	Q9CV72_MOUSE	Q9CV72_mus musculus
7	35	81.4	300	Q8C1Z6_MOUSE	Q8C1Z6_mus musculus
8	35	81.4	359	Q8C886_MOUSE	Q8C886_mus musculus
9	35	81.4	462	Q9ADR7_STAAU	Q9ADR7_staphylococ
10	35	81.4	503	Q414S5_GIBZE	Q414S5_gibberella
11	35	81.4	513	Q75BS7_ASHGO	Q75BS7_ashbya goss
12	35	81.4	552	K2C6A_MOUSE	P50446_mus musculus
13	35	81.4	554	Q3UV11_MOUSE	Q3UV11_mus musculus
14	35	81.4	561	K2C6B_MOUSE	Q92331_mus musculus
15	35	81.4	581	Q61FZ5_RAT	Q61FZ5_rattus norv
16	35	81.4	594	Q3UV17_MOUSE	Q3UV17_mus musculus
17	35	81.4	629	K2C3_HUMAN	P12035_homo sapien
18	35	81.4	629	K2C3_RABIT	Q29426_oryctolagus
19	35	81.4	636	Q53655_STAAU	Q53655_staphylococ
20	35	81.4	636	Q5HJ99_STAAU	Q5HJ99_staphylococ
21	35	81.4	690	Q8RUNK_STAAU	Q8RUNK_staphylococ
22	35	81.4	711	Q8BPM4_MOUSE	Q8BPM4_mus musculus
23	35	81.4	729	Q55545_SYNY3	Q55545_synecocyst
24	35	81.4	1136	Q8C2N6_MOUSE	Q8C2N6_mus musculus
25	35	81.4	1165	Q6PG61_MOUSE	Q6PG61_mus musculus
26	35	81.4	1217	Q5DTM5_MOUSE	Q5DTM5_mus musculus
27	35	81.4	1979	Q88196_MOUSE	Q88196_mus musculus
28	35	81.4	3774	Q2U2A1_ASPOR	Q2U2A1_aspergillus
29	35	81.4	4186	Q4ZVR9_PSEU2	Q4ZVR9_pseudomonas
30	35	81.4	4190	Q83Y48_PSEEX	Q83Y48_pseudomonas
31	34	79.1	93	Q6F7V2_ACIAO	Q6F7V2_acinetobact

32	34	79.1	94	2	Q7VW78 BORPE	Q7VW78 bordetella
33	34	79.1	94	2	Q7W5N8 BORPA	Q7W5N8 bordetella
34	34	79.1	94	2	Q7WD74 BORBR	Q7WD74 bordetella
35	34	79.1	228	2	Q898V5_CLOTE	Q898V5 clostridium
36	34	79.1	532	2	Q75DG5_ASHGO	Q75DG5 ashbya goss
37	34	79.1	670	1	REP_BUGBP	Q89A21 buchnera ap
38	34	79.1	703	2	Q7NBD8_MYCGA	Q7NBD8 mycoplasma
39	34	79.1	730	2	Q7RUA4_NEUCR	Q7RUA4 neurospora
40	34	79.1	740	2	Q6W901_NEUCR	Q6W901 neurospora
41	34	79.1	823	2	Q2NKG6_ECOLI	Q2NKG6 escherichia
42	34	79.1	823	2	O50330_ECOLI	O50330 escherichia
43	34	79.1	950	2	Q3P2E1_NITHA	Q3P2E1 nitrobacter
44	34	79.1	1009	2	Q2RQ71_RHORU	Q2RQ71 rhodospirill
45	34	79.1	1101	2	Q50XY2_ENTHI	Q50XY2 entamosba h

## ALIGNMENTS

RESULT 1  
Q40275\_KINRA PRELIMINARY; PRT; 587 AA.  
AC Q40275;  
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.  
DT 27-SEP-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Similar to Coenzyme F420-dependent N5 N10-methylene  
DE tetrahydromethanopterin reductase and related flavin-dependent  
DE oxidoreductases.  
GN ORFNames=KradDRAFT 1489;  
OS Kineococcus radiotolerans SRS30216.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Frankineae; Kineosporiaceae; Kineococcus.  
OX NCBI\_TaxID=266940;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-SRS30216;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,  
RA Hammon N., Istrati S., Pitluck S., Richardson P.,  
RT "Sequencing of the draft genome and assembly of Kineococcus  
RT radiotolerans SRS30216";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-SRS30216;  
RG US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.,  
RT "Annotation of draft genome assembly of Kineococcus radiotolerans  
RT SRS30216";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-SRS30216;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,  
RA Hammon N., Istrati S., Pitluck S., Richardson P.,  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial License  
CC  
CC EMBL; AAEP02000040; EAM74395.1; -; Genomic\_DNA.  
DR InterPro; IPR011251; Luciferase like  
SQ SEQUENCE 587 AA; 64579 MW; 86DB19E29DCFE925 CRC64;

Query Match 90.7%; Score 39; DB 2; Length 587;  
Best Local Similarity 87.5%; Pred. No. 80;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 AQYEDYQAQ 8
Db      422 AQYEDYQAQ 429

RESULT 2
Q35ANO_9BRAD      PRELIMINARY; PRT; 948 AA.
AC Q35ANO;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE Transaldolase subfamily.
GN ORNames=BrADRAFT_1619;
OS Bradyrhizobium sp. BTail.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=288000;
RN [1]_SEQUENCE 288000;

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BTail;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Bradyrhizobium sp
RT BTail";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BTail;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Bradyrhizobium sp. BTail.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALJ01000012; EAP27068.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004347; F:glucose-6-phosphate isomerase activity; IEA.
DR GO; GO:0004801; F:transaldolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006094; P:gluconeogenesis; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR GO; GO:0006098; P:pentose-phosphate shunt; IEA.
SQ SEQUENCE 948 AA; 102058 MW; 02AB526AE5BC5378 CRC64;

Query Match 88.4%; Score 38; DB 2; Length 948;
Best Local Similarity 87.5%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AQYEDYQAQ 8
Db      439 AQYEDYQAQ 446

RESULT 3
Q8FRV4_COREF      PRELIMINARY; PRT; 212 AA.
AC Q8FRV4;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein.
GN OrderedLocNames=CE0654;
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;

```

```

RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Y8-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;
RA Nishio Y., Nakamura Y., Kawarabayashi Y., Ueda Y., Kimura E.,
RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA Gojobori T.;
RT "Comparative complete genome sequence analysis of the amino acid
RT replacements responsible for the thermostability of Corynebacterium
RT efficiens.";
RL Genome Res. 13:1572-1579 (2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BA00035; BAC17464.1; -; Genomic_DNA.
DR Biocyc; CEFF196164; CE0654-MONOMER; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 212 AA; 22515 MW; EFDD9C365DC0E0E5 CRC64;

Query Match 86.0%; Score 37; DB 2; Length 212;
Best Local Similarity 87.5%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AQYEDYQAQ 8
Db      171 AQYEDYQAQ 178

RESULT 4
Q9IB36_KATPE      PRELIMINARY; PRT; 198 AA.
ID Q9IB36_KATPE
AC Q9IB36;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 07-FEB-2006, entry version 23.
DE Myosin light chain 1.
GN Name=mlc1;
OS Katsuwonus pelamis (Skipjack tuna) (Bonito).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Scombroidei;
OC Scombridae; Katsuwonus.
OX NCBI_TaxID=8226;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ishizaki S., Masuda Y., Tanaka M., Hirayama Y., Watabe S.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB042037; BAA95126.1; -; mRNA.
DR HSSP; P14649; 1OE9.
DR SMR; Q9IB36; 1-279, 139-347, 373-427.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR011992; EF-Hand type.
DR InterPro; IPR002048; EF-hand_Ca_bd.
DR Pfam; PF00036; ehand; 2.
DR ProDom; PD000012; EF-hand; 2.
DR PROSITE; PS50222; EF_HAND_2; 2.
KW Repeat.
SQ SEQUENCE 198 AA; 21349 MW; 31D1BC98D780D46C CRC64;

Query Match 83.7%; Score 36; DB 2; Length 198;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 AQYEDYQAQ 8
Db      131 AQYEDYQAQ 138

```



```

RESULT 5
Q3KDJ2_PSEPF
AC Q3KDJ2_PSEPF PRELIMINARY; PRT; 210 AA.
ID Q3KDJ2;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 21-FEB-2006, entry version 5.
DE Glutathione S-transferase-like.
GN OrderedLocNames=Pf1_2422;
OS Pseudomonas fluorescens (strain Pfo-1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205922;
RN [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP US DOE Joint Genome Institute;
RG Copeland A., Lucas S., Litlick S., Saunders E.H., Schmutz J.,
RA Hammon N., Istanti S., Pitluck S., Land M., Kyriakides N., Richardson P.;
RA Larimer F., Land M., Kyriakides N., Richardson P.;
RT "Complete sequence of Pseudomonas fluorescens Pfo-1."
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the GST superfamily.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; CP000094; ABA74163.1; -; Genomic DNA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR004045; GST_N.
DR InterPro; IPR012336; Thioredoxin-like fd.
DR InterPro; IPR012335; Thioredoxin_fold.
DR Pfam; PF02798; GST_N; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 210 AA; 24148 MW; FDE66BC52C7BC9E9 CRC64;

Query Match 81.4%; Score 35; DB 2; Length 210;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQVEDYQAQ 8
DB 184 AQLEDYQAQ 191

RESULT 6
Q9CV72_MOUSE
ID Q9CV72_MOUSE PRELIMINARY; PRT; 278 AA.
AC Q9CV72;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 22.
DE Adult male tongue cDNA, RIKEN full-length enriched library,
DE clone:231008N03 product:keratin complex 2, basic, gene 6b, full
DE insert sequence. (Fragment).
GN Name=Krt2-6b;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schraml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,

```



RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nieuwen E., Vexard R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashina T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.,  
RA "The transcriptional landscape of the mammalian genome.",  
RL Science 309:1559-1563 (2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome";  
RL Science 309:1564-1566 (2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Clothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Malcais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verdaro R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner T., Bono H., Kasukawa T., Saito R.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690 (2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630 (2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Yamawake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771 (2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial license  
CC -----  
CC EMBL; AK089974; BAC41023.1; -; mRNA.  
DR HSSP; P53041; 1A17.  
DR Ensembl; ENSMUSG00000040785; Mus musculus.  
DR MGI; MGI:1276539; Ttc3.  
DR InterPro; IPR013026; TPR\_region.  
DR PROSITE; PS50005; TPR; 1.  
SQ PROSITE 300 AA; 35253 MW; D6F6AFCD4CA19EDA CRC64;  
Query Match 81.4%; Score 35; DB 2; Length 300;  
Best Local Similarity 85.7%; Pred No. 2.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 QYEDYQAQ 8  
Db 51 QYKDYQAQ 57  
||:||||  
||:||||  
RESULT 8  
Q8C886 MOUSE PRELIMINARY; PRT; 359 AA.  
ID Q8C886 MOUSE  
AC Q8C886  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-FEB-2006, entry version 16.  
DE 16 days embryo head cDNA, RIKEN full-length enriched library,  
DE clone: C130033N06 cDNA; inferred: hypothetical protein (Homo  
DE sapiens), full insert sequence.  
DE Name=Plekhn1,  
GN

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Rodentia; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chaik A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Humnicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasaki Y., Kelson J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone P., Miyake S., Morris K.,  
 RA Mottagui-Fabriz S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Roat B., Ruan Y., Salzberg S.L., Sandelin A., Seno S., Sessa L., Sheng Y.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugitara K., Sultana R., Takenaka Y., Taki K.,  
 RA Tamajima K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kal C., Sakai D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RT (Genome Network Core Team) and the FANTOM Consortium;  
 RC "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,

Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 GAasterland T., Gariboldi M., Giasi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sakai D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitesunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Mateumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Oawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Head;

```

RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saiko D., Saichon H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AK048084; BAC33236.1; -; mRNA.
DR MGI; MGI:2387630; Plekhl1.
DR InterPro; IPR011993; PH_type.
KW Hypothetical protein.
SQ SEQUENCE 359 AA; 39358 MW; F42CF809DFDAE7DD CRC64;

Query Match 81.4%; Score 35; DB 2; Length 359;
Best Local Similarity 75.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8
Db 117 ASYEDYSQ 124

RESULT 9
Q9ADR7 STAAU PRELIMINARY; PRT; 462 AA.
AC Q9ADR7;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Coagulase (fragment).
GN Name=coa;
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=mrsea;
RA Carter P.E., Begbie K., Thomson-Carter F.M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AJ309189; CAC36440.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 462 462
SQ SEQUENCE 462 AA; 52952 MW; 1C9BDA535AC86238 CRC64;

Query Match 81.4%; Score 35; DB 2; Length 462;
Best Local Similarity 75.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8
Db 77 SQYEDYKQ 84

RESULT 10
Q414S5 GIBZE
ID Q414S5 GIBZE PRELIMINARY; PRT; 503 AA.
AC Q414S5;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.

```

```

DE Hypothetical protein.
GN ORFNames=FG07783.1;
OS Gibberella zeae (Fusarium graminearum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocrymvetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5518;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=PH-1 / NRRL 31084;
RA Birren B.W., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Chospel Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D.,
RA Galagan J.E., Gardyna S., Gnerre S., Graham L., Grand-Pierre N.,
RA Hafez I., Hagopian D., Hagos B., Hall J., Horton L., Hulme W.,
RA Iliev I., Jaffe D., Johnson R., Jones C., Kamal M., Kamat A.,
RA Karatas A., Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G.,
RA Lui A., Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J.,
RA Manning J., Matthews C., Mauceli E., McCarthy M., Meldrim J.,
RA Menes L., Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C.,
RA Nicol R., Nielsen C.B., Norbu C., O'Connor T., O'Donnell P.,
RA O'Neill D., Oliver J., Peterson K., Phunkhang P., Pierre N.,
RA Purcell S., Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C.,
RA Rogov P., Roman J., Schauer S., Schupback R., Seaman S., Severy P.,
RA Smirnov S., Smith C., Spencer B., Stange-Thomann N., Stojanovic N.,
RA Stubbs M., Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.S.;
RT "Fusarium graminearum genome sequence.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AACM01000320; EAA7977.1; -; Genomic DNA.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 503 AA; 56594 MW; 86EDA10DS0C8D473 CRC64;

Query Match 81.4%; Score 35; DB 2; Length 503;
Best Local Similarity 75.0%; Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8
Db 118 AEYEDYIQ 125

RESULT 11
Q75BS7 ASHCO PRELIMINARY; PRT; 513 AA.
ID Q75BS7 ASHCO PRELIMINARY; PRT; 513 AA.
AC Q75BS7;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE ACR194Cp.
GN Ordered locusNames=ACR194C;
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 10895;
RX PubMed=15001715; DOI=10.1126/science.1095781;
RA Dietrich F.S., Voegelé S., Brachat S., Lerch A., Gates K., Steiner S.,
RA Mohr C., Pöhlmann R., Luedi P., Choi S., Wing R.A., Flavler A.,

```

```

RA Gaffney T.D., Philippen P.;
RT "The Ashbya gossypii genome as a tool for mapping the ancient
RT Saccharomyces cerevisiae genome.";
RL Science 304:304-307(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AE016816; AAS51420.1; -; Genomic_DNA.
DR AGD; ACRI94C; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR PRINTS; PR00320; GPROTEINERPT.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Complete proteome; Repeat; WD repeat.
SQ SEQUENCE 513 AA; 57793 MW; 6B92DD1938891760 CRC64;

Query Match 81.4%; Score 35; DB 2; Length 513;
Best Local Similarity 85.7%; Pred. No. 4.1e+02; Gaps 0;
Matches 6; Conservative 1; Mismatches 0; Indels 0;

Qy 1 AQYEDYA 7
Db 371 SQYEDYA 377

RESULT 12
K2C6A MOUSE
ID K2C6A MOUSE STANDARD; PRT; 552 AA.
AC P50446; O9Z332;
DT 01-OCT-1986, integrated into UniProtKB/Swiss-Prot.
DT 01-MAR-2005, sequence version 2.
DT 21-FEB-2006, entry version 42.
DE Keratin, type II cytoskeletal 6A (Cytokeratin-6A) (CK 6A) (K6a
DE keratin) (Keratin-6 alpha) (mk6-alpha).
GN Name=Krt6a; Synonyms=Ker2, Krt2-6, Krt2-6a, Krt6;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN NUCLEOTIDE SEQUENCE [MRNA].
RC TISSUE=Epidermis;
RX MEDLINE=85014838; PubMed=6207530;
RA Steinert P.M., Parry D.A.D., Racoosin E.L., Idler W.W., Steven A.C.,
RA Trus B.L., Roop D.R.;
RA "The complete cDNA and deduced amino acid sequence of a type II mouse
RT epidermal keratin of 60,000 Da: analysis of sequence differences
RT between type I and type II keratins.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5709-5713(1984).
[2]
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA], TISSUE SPECIFICITY, AND INDUCTION.
RC STRAIN=129/Sv; TISSUE=Skin;
RX MEDLINE=99009332; PubMed=9790766; DOI=10.1006/geno.1998.5476;
RA Takahashi K., Yan B., Yamanishi K., Imanura S., Coulombe P.A.;
RT "The two functional keratin 6 genes of mouse are differentially
RT regulated and evolved independently from their human orthologs.";
RL Genomics 53:170-183(1998).
[3]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Jaw, and Limb;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore H., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,

```

```

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.B.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RN NUCLEOTIDE SEQUENCE OF 527-552.
RX PubMed=1713533;
RA Finch J., Andrews K., Krieg P., Furstenberger G., Slaga T.,
RA Otsuyama A., Tanooka H., Bowden G.T.;
RT "Identification of a cloned sequence activated during multi-stage
RT carcinogenesis in mouse skin.";
RL Carcinogenesis 12:1519-1522(1991).
CC -!- FUNCTION: There are two types of cytoskeletal and microfibrillar
CC keratin: I (acidic; 40-55 kDa) [K9 to K20] and II (neutral to
CC basic; 56-70 kDa) [K1 to K8]. Both a basic and an acidic keratin
CC are required for filament assembly.
CC -!- SUBUNIT: Heterodimer of a type I and a type II keratin. KRT6
CC isomers associate with KRT16 and/or KRT17.
CC -!- TISSUE SPECIFICITY: Predominates in the adult trunk skin, tongue,
CC trachea/esophagus and eye. In adult skin, localization is
CC restricted to hair follicles, where it is localized predominantly
CC in the outer root sheath.
CC -!- INDUCTION: With the exception of specific body sites, keratin 6
CC expression is induced under conditions of epithelial
CC hyperproliferation such as wound healing, certain skin diseases,
CC cancer, and by treatment of the skin with the phorbol ester PMA.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; K02108; AAA39395.1; -; mRNA.
CC EMBL; AB012033; BAA34178.1; -; Genomic_DNA.
CC EMBL; BC080820; AAH80820.1; -; mRNA.
CC PTR; I59009; I59009.
CC HSP; P08670; LGK7.
CC IntAct; P50446; -.
CC Ensembl; ENSMUSG0000058354; Mus musculus.
CC MGI; MGI:1100845; Krt2-6a.
CC GO; GO:0005882; C:intermediate filament; ISS.
CC GO; GO:0005198; F:structural molecule activity; ISS.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC InterPro; IPR001664; IF.
CC InterPro; IPR003054; Keratin_II.
CC PANTHER; PTHR1893; SFS; Keratin_II; 1.
CC Pfam; PF00038; Filament; 1.
CC PRINTS; PR01276; TYPE2KERATIN.
CC PROSITE; PS00226; IF; 1.
CC Coiled coil; Intermediate filament; Keratin; Phosphorylation.
FT INIT_MET 0 0 By similarity.
FT CHAIN 1 552 Keratin, type II cytoskeletal 6A.
FT /Frid=PRO_0000063736.
FT REGION 1 150 Head.
FT REGION 151 460 Rod.
FT REGION 151 186 Coiled coil 1A.
FT REGION 187 205 Coiled coil 1.
FT REGION 206 297 Coiled coil 1B.
FT REGION 298 321 Coiled coil 12.
FT REGION 322 460 Coiled coil 2.
FT REGION 461 552 Tail.
FT SITE 402 402 Stutter.
FT MOD_RES 57 57 Phosphoserine (By similarity).
FT CONFLICT 23 23 P -> L (in Ref. 1).
FT CONFLICT 120 120 P -> L (in Ref. 1).

```



```

FT CONFLICT 172 172 L -> M (in Ref. 1).
FT CONFLICT 179 179 L -> M (in Ref. 1).
FT CONFLICT 190 190 G -> D (in Ref. 1).
FT CONFLICT 198 198 L -> M (in Ref. 1).
FT CONFLICT 223 224 LD -> MN (in Ref. 1).
FT CONFLICT 232 239 DTVEDYKS -> ELVEELRN (in Ref. 1).
FT CONFLICT 250 250 A -> D (in Ref. 1).
FT CONFLICT 311 311 D -> V (in Ref. 1).
FT CONFLICT 317 317 D -> V (in Ref. 1).
FT CONFLICT 329 331 YED -> FEV (in Ref. 1).
FT CONFLICT 343 343 W -> L (in Ref. 1).
FT CONFLICT 431 431 R -> M (in Ref. 1).
FT CONFLICT 437 437 Q -> H (in Ref. 1).
FT CONFLICT 497 497 L -> M (in Ref. 1).
SQ SEQUENCE 552 AA; 59204 MW; DEB2DA645EBA8DAF CRC64;

Query Match 81.4%; Score 35; DB 1; Length 552;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDYQAQ 8
Db 327 AQVEDYQAQ 334

RESULT 13
Q3UV11_MOUSE PRELIMINARY; PRT; 554 AA.
AC Q3UV11;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Adult female vagina cDNA, RIKEN full-length enriched library,
DE clone:9300330J10 product:keratin complex 2, basic, gene 6B, full
DE insert sequence.
GN Name=Krt2-6b;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Vagina;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Vagina;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Imbondato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhury V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Heminger P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi S., Hensch T.K., Hirokawa N.,
RA Hill D., Humnick L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,

```

```

RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Suglura K., Sultana R., Takenaka Y., Taki K.,
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimonard S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Vagina;
RX PubMed=16141073; DOI=10.1126/science.1112009;
RG RIKEN Genome Exploration Research Group, and Genome Science Group
RG (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Vagina;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimonard S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszewski-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Vagina;
RX MEDLINE=21085560; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Iwama M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fetschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

```

```
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann J., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.,
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RN [6]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Vagina; DOI=10.1101/gr.145100;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayase N., Sugahara Y., Shibata K., Itoh M.,
RA Kanno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [7]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Vagina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kanno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yanamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Oyawa Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [8]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Vagina;
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,
RA Kawai J., Kojima M., Kono H., Murata M., Nakamura M., Ninomiya N.,
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
RA Muramatsu M., Hayashizaki Y.;
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RL Nucleotide sequence data entry.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK137688; BAE23462.1; -; mRNA.
DR MGI; MGI:1333768; Krt2-6b.
DR GO; GO:0016021; C:integral to membrane; RCA.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR GO; GO:0045109; P:intermediate filament organization; IGI.
DR GO; GO:0031424; P:keratinization; IGI.
DR GO; GO:0002009; P:morphogenesis of an epithelium; IGI.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
DR Intermediate filament; Keratin.
SQ SEQUENCE 554 AA; 59526 MW; 440DE8AE3984B8AA CRC64;
Query Match 81.4%; Score 35; DB 2; Length 554;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AQYEDYAQ 8
D5 328 AQYEDIAQ 335
*RESULT 14
```

```
K2C6B MOUSE
ID K2C6B_MOUSE STANDARD; PRT; 561 AA.
AC Q92331;
DT 15-FEB-2005, integrated into UniProtKB/Swiss-Prot.
DT 15-FEB-2005, sequence version 2.
DT 07-FEB-2006, entry version 33.
DE Keratin, type II cytoskeletal 6B (Cytokeratin-6B) (CK 6B) (K6b
DE keratin) (Keratin-6 beta) (mk6-beta).
GN Name=krt6b; Synonyms=k6-beta, Krt2-6b;
OS Mus musculus (Mouse).
OC Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Muroidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA], TISSUE SPECIFICITY, AND INDUCTION.
RP STRAIN=129/SV; TISSUE=Skin;
RX MEDLINE=99009332; PubMed=9790766; DOI=10.1006/geno.1998.5476;
RA Takahashi K., Yan B., Yamashita K., Imamura S., Coulombe P.A.;
RT "The two functional keratin 6 genes of mouse are differentially
RT regulated and evolved independently from their human orthologs.";
RL Genomics 53:170-183(1998).
CC -!- FUNCTION: There are two types of cytoskeletal and microfibrillar
CC keratin: I (acidic; 40-55 kDa) [K9 to K20] and II (neutral to
CC basic; 56-70 kDa) [K1 to K8]. Both a basic and an acidic keratin
CC are required for filament assembly.
CC -!- SUBUNIT: Heterodimer of a type I and a type II keratin. KRT6
CC isomers associate with KRT16 and/or KRT17.
CC -!- TISSUE SPECIFICITY: Expressed in adult epithelia including the
CC tongue, esophagus/trachea, eye and skin. Localized preferentially
CC to the suprabasal layers of thickened epidermis in injured and
CC chemically treated skin.
CC -!- INDUCTION: By injury, and treatment of the skin with the phorbol
CC ester PMA.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB012042; BAA34179.1; -; Genomic_DNA.
DR HSRP; P08670; IGK7.
DR MGI; MGI:1333768; Krt2-6b.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0005882; C:intermediate filament; ISS.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR GO; GO:0005198; F:structural molecule activity; ISS.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR18893; SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
DR Coiled coil; Intermediate filament; Keratin; Phosphorylation.
FT INIT MET 0 0 By similarity.
FT CHAIN 1 561 Keratin, type II cytoskeletal 6B.
FT REGION 1 158 Head.
FT REGION 159 468 Rod.
FT REGION 159 194 Coiled 1A.
FT REGION 195 213 Linker 1.
FT REGION 214 305 Coiled 1B.
FT REGION 306 329 Linker 12.
FT REGION 330 468 Coiled 2.
FT REGION 469 561 Tail.
FT SITE 410 410 Stutter.
FT MOD RES 57 57 Phosphoserine (By similarity).
SQ SEQUENCE 561 AA; 60191 MW; 65B71CAB7F9564E1 CRC64;
Query Match 81.4%; Score 35; DB 1; Length 561;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```



OY 1 AQYEDYAO 8  
|||||  
Db 335 AQYEDIAQ 342

RESULT 15  
Q6IFZ5 RAT  
ID Q6IFZ5 RAT PRELIMINARY; PRT; 581 AA.  
AC Q6IFZ5;  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 13.  
DE Type II keratin KB9.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;  
RA Hesse M., Zimek A., Weber K., Magin T.M.;  
RT "Comprehensive analysis of keratin gene clusters in humans and  
rodents.";  
RL Eur. J. Cell Biol. 83:19-26(2004).  
CC -!- MISCELLANEOUS: The sequence shown here is derived from an  
EMBL/GenBank/DBJ third party annotation (TPA) entry.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BK003994; DAA02235.1; -; mRNA.  
DR Ensembl; ENSRNOG00000031785; Rattus norvegicus.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin I.  
DR InterPro; IPR003054; Keratin II.  
DR PANTHER; PTHR18893.SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 581 AA; 61758 MW; 13B0F3D23BDF4E49 CRC64;

Query Match 81.4%; Score 35; DB 2; Length 581;  
Best Local Similarity 87.5%; Pred. NO. 4.7e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQYEDYAO 8  
|||||  
Db 337 AQYEDIAQ 344

Search completed: July 5, 2006, 19:01:29  
Job time : 83.9641 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 69.5146 Seconds  
(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928a-10  
Perfect score: 43  
Sequence: 1 AQYEDYQA 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_8.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*  
10: Geneseq2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	43	100.0	8	AAO26390	Aao26390 Psoriasis
2	43	100.0	8	ADV04415	Adv04415 Leishmani
3	35	81.4	9	AAO26382	Aao26382 Psoriasis
4	35	81.4	9	ADV04407	Adv04407 Leishmani
5	35	81.4	30	AAV14568	Aav14568 Partial a
6	35	81.4	420	ADP29879	Adp29879 Human sec
7	35	81.4	541	AAU75491	Aau75491 S. aureus
8	35	81.4	629	AEA15501	Aea15501 Human pol
9	35	81.4	636	AAU34342	Aau34342 Staphyloc
10	35	81.4	636	AAU37415	Aau37415 Staphyloc
11	35	81.4	636	ADA89509	Ada89509 Staphyloc
12	35	81.4	636	ABJ18985	Abj18985 Pathogen
13	35	81.4	636	ABM72027	Abm72027 Staphyloc
14	35	81.4	644	AEA15501	Aea15501 Human pol
15	34	79.1	315	ADF06590	Adf06590 Bacterial
16	34	79.1	661	AAE63226	Aae63226 Mycoplasma
17	34	79.1	661	AAE79910	Aae79910 M.gallise
18	33	76.7	688	AAE07559	Aae07559 Protein e
19	33	76.7	1178	ABM70281	Abm70281 Photorhab
20	33	76.7	2011	ABM70279	Abm70279 Photorhab
21	33	76.7	3241	ABU22000	Abu22000 Protein e
22	32	74.4	176	ABB67136	Abb67136 Drosophil
23	32	74.4	252	ABE41195	Abe41195 L. pneumo

24	32	74.4	253	9	ABE37890	Aeb37890 L. pneumo
25	32	74.4	587	4	AAU34198	Aau34198 Staphyloc
26	32	74.4	603	4	AAU37074	Aau37074 Staphyloc
27	32	74.4	603	6	ABU16392	Abu16392 Protein e
28	32	74.4	603	6	ABM72619	Abm72619 Staphyloc
29	31	72.1	11	5	ABG73045	Abg73045 Cryptococ
30	31	72.1	49	8	ADO39763	Ado39763 Caenorhab
31	31	72.1	82	4	AAW20776	Aaw20776 Peptide #
32	31	72.1	82	4	ABB42330	Abb42330 Peptide #
33	31	72.1	82	4	AAW36136	Aaw36136 Peptide #
34	31	72.1	82	4	ABB25821	Abb25821 Protein #
35	31	72.1	82	4	AAW76029	Aaw76029 Human bon
36	31	72.1	82	4	AAW63216	Aaw63216 Human bra
37	31	72.1	82	4	ABG57755	Abg57755 Human liv
38	31	72.1	82	5	ABG45427	Abg45427 Human pep
39	31	72.1	106	4	AAU27962	Aau27962 Human con
40	31	72.1	125	7	ADH86301	Adh86301 Enterococ
41	31	72.1	131	4	AAW82389	Aaw82389 S. epider
42	31	72.1	134	9	ADX58875	Adx58875 Human pig
43	31	72.1	150	5	ABP38873	Abp38873 Staphyloc
44	31	72.1	150	8	ADS06504	Ads06504 Staphyloc
45	31	72.1	179	4	ABB62918	Abb62918 Drosophil

## ALIGNMENTS

RESULT 1  
AAO26390  
ID AAO26390 standard; peptide; 8 AA.  
XX  
AC AAO26390;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 10.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
(AKIV-) AKIVA LLC.  
XX  
O'daly JA;  
XX  
DR WPI; 2003-018763/01.  
XX  
New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
Claim 1; Page 43; 56pp; English.  
XX  
The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention

## SQ Sequence 8 AA;

Query Match 100.0%; Score 43; DB 6; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDYQAQ 8  
| | | | |  
Db 1 AQVEDYQAQ 8

## RESULT 2

ADV04415  
ID ADV04415 standard; peptide; 8 AA.

XX AC ADV04415;

XX DT 24-FEB-2005 (first entry)

XX DE Leishmania immunotherapeutic peptide SEQ ID NO:10.

XX KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;  
XX KW psoriasis.

XX OS Leishmania sp.

XX PN US2004241168-A1.

XX PD 02-DEC-2004.

XX PF 09-FEB-2004; 2004US-00774928.

XX PR 16-MAR-2001; 2001US-00809003.

XX FR 17-OCT-2003; 2003US-00687892.

XX FA (ODAL/) ODALY J A.

XX PI Odaly JA;

XX DR WPI; 2005-011563/01.

XX PT Inhibiting selectively T-cell rolling in human, by administering compound  
interfering with cutaneous lymphocyte antigen-E selectin interaction and  
leukocyte function associated antigen-1 interaction.

XX PS Claim 18; SEQ ID NO 10; 21pp; English.

XX CC The invention relates to a novel method for selectively inhibiting T-cell  
rolling in a human host, comprising administering a compound that  
selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
selectin interaction and leukocyte function associated antigen (LFA)-  
1/intercellular adhesion molecule (ICAM) and very late antigen  
(VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
includes an immunotherapeutic agent, which comprises a purified protein  
extract that is isolated by diethylaminoethyl Sephadex chromatography of  
a Nontidet P-40 insoluble particulate antigen fraction derived from  
isolated killed cells of anastigotes from one or more species of the  
Leishmania genus, where the particulate antigen fraction is solubilized  
with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
to diethylaminoethyl Sephadex and is eluted with a solution comprising  
0.1 M sodium chloride, 8 M urea and 0.025 M  
Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract  
includes polypeptides having apparent molecular weights of 73, 80 and 82  
kDa, after total reduction and alkylation. The species is L. amazonensis,  
L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the  
invention has antipsoriatic and immunostimulant activity, and inhibits T-  
cell rolling by interfering with CLA-E selectin interaction and LFA-  
1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
inhibiting T-cell rolling in a human host, and for treating psoriasis.  
XX CC The present sequence represents a Leishmania peptide used in the  
invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 43; DB 9; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQVEDYQAQ 8  
| | | | |  
Db 1 AQVEDYQAQ 8

## RESULT 3

AAO26382  
ID AAO26382 standard; peptide; 9 AA.

XX AC AAO26382;

XX DT 30-JAN-2003 (first entry)

XX DE Psoriasis treating immunotherapeutic peptide, SEQ ID NO 2.

XX KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
XX KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
XX KW gene therapy.

XX OS Leishmania sp.

XX PN WO200274239-A2.

XX PD 26-SEP-2002.

XX PF 04-MAR-2002; 2002WO-US006496.

XX PR 16-MAR-2001; 2001US-00809003.

XX PA (AKIV-) AKIVA LLC.

XX PI O'daly JA;

XX DR WPI; 2003-018763/01.

XX PT New polypeptide and immunogenic variants comprising amino acid sequences  
of particulate antigens, useful for the treatment and clinical remission  
of psoriasis.

XX PS Claim 1; Page 43; 56pp; English.

XX CC The invention relates to a polypeptide comprising an isolated amino acid  
sequence or immunogenic variants selected from any of 14 fully defined  
sequences of 7-16 amino acids, given in the specification. The  
CC immunotherapeutic agents and a pharmaceutical compositions comprising  
CC polynucleotides and vectors of the invention are useful for the treatment  
CC and clinical remission of psoriasis. The isolated nucleic acids are  
CC useful as probes. The sequences of the invention can be used in the  
CC treatment of disorders by gene therapy. This sequence represents one of  
the 14 immunotherapeutic peptides of the invention

XX SQ Sequence 9 AA;

Query Match 81.4%; Score 35; DB 6; Length 9;  
Best Local Similarity 87.5%; Pred. No. 2.1e+06;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQVEDYQAQ 8  
| | | | |  
Db 1 AQVEDYQAQ 8

## RESULT 4

ADV04407  
ID ADV04407 standard; peptide; 9 AA.

XX AC ADV04407;

XX XX

DT 24-FEB-2005 (first entry)  
XX Leishmania immunotherapeutic peptide SEQ ID NO:2.  
DE  
XX  
XX  
KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;  
KW psoriasis.  
XX  
OS Leishmania sp.  
XX  
PN US2004241168-A1.  
XX  
PD 02-DEC-2004.  
XX  
XX 09-FEB-2004; 2004US-00774928.  
PF  
XX 16-MAR-2001; 2001US-00809003.  
XX  
PR 17-OCT-2003; 2003US-00687892.  
XX  
PA (ODAL/) ODALY J A.  
XX  
PI Odaly JA;  
XX  
XX WPI; 2005-011563/01.  
DR  
XX Inhibiting selectively T-cell rolling in human, by administering compound  
PT interfering with cutaneous lymphocyte antigen-E selectin interaction and  
PT leukocyte function associated antigen-1 interaction.  
XX  
PS Claim 9; SEQ ID NO 2; 21pp; English.  
XX  
XX The invention relates to a novel method for selectively inhibiting T-cell  
CC rolling in a human host, comprising administering a compound that  
CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
CC selectin interaction and leukocyte function associated antigen (LFA)-  
CC 1/intercellular adhesion molecule (ICAM) and very late antigen  
CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
CC includes an immunotherapeutic agent, which comprises a purified protein  
CC extract that is isolated by diethylaminoethyl Sephadex chromatography of  
CC a Nonidet P-40 insoluble particulate antigen fraction derived from  
CC isolated killed cells of amastigotes from one or more species of the  
CC Leishmania genus, where the particulate antigen fraction is solubilized  
CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
CC to diethylaminoethyl Sephadex and is eluted with a solution comprising  
CC 0.1 M sodium chloride, 8 M urea and 0.025 M  
CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract  
CC includes polypeptides having apparent molecular weights of 73, 80 and 82  
CC kDa, after total reduction and alkylation. The species is L. amazonensis,  
CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the  
CC invention has antipsoriatic and immunostimulant activity, and inhibits T-  
CC cell rolling by interfering with CLA-E selectin interaction and LFA-  
CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
CC inhibiting T-cell rolling in a human host, and for treating psoriasis.  
CC The present sequence represents an Leishmania peptide used in the  
CC invention.  
XX  
XX  
SQ Sequence 9 AA;  
Query Match 81.4%; Score 35; DB 9; Length 9;  
Best Local Similarity 87.5%; Pred. No. 2.1e+06;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AQVEDYQAQ 8  
Db 1 AQVEDIAQ 8  
RESULT 5  
ID AAY14568  
XX AAY14568 standard; peptide; 30 AA.  
XX  
AC AAY14568;  
XX  
DT 14-SEP-1999 (first entry)

XX Partial amino acid sequence of D-sequence binding protein.  
DE  
XX  
XX D-sequence; binding protein; phosphorylation; adeno-associated virus;  
KW AAV; transgene; expression; DNA synthesis; cancer; EMSA; detection;  
KW competitive electrophoretic mobility shift assay.  
XX  
OS Synthetic.  
XX  
PN WO9911652-A1.  
XX  
PD 11-MAR-1999.  
XX  
XX 01-SEP-1998; 98WO-US018194.  
PF  
XX 02-SEP-1997; 97US-0056052P.  
PR  
XX (ADRE-) ADVANCED RES & TECHNOLOGY INST.  
PA  
XX Srivastava A, Qing K, Wang X, Ponnazhagan S, Bajpai A;  
PI WPI; 1999-394757/33.  
XX  
XX New isolated D sequence binding protein.  
PT  
XX Disclosure; Page 23; 163pp; English.  
PS  
XX The invention relates to an isolated and purified D sequence binding  
CC protein (D-BP) having a molecular weight of about 53 kD. This peptide  
CC represents a partial amino acid sequence of the complete D-BP. The D-BP  
CC is phosphorylated at tyrosine residues and blocks AAV-mediated transgene  
CC expression in infected cells by inhibiting the leading strand viral DNA  
CC synthesis. Inhibiting the function of D-BP in host cells can increase  
CC expression of selected nucleic acids from an AAV, particularly for  
CC increasing therapeutic nucleic acids for treating a disease such as  
CC cancer.  
XX  
SQ Sequence 30 AA;  
Query Match 81.4%; Score 35; DB 2; Length 30;  
Best Local Similarity 87.5%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AQVEDYQAQ 8  
Db 17 AQVEDIAQ 24  
RESULT 6  
ADP29879  
ID ADP29879 standard; protein; 420 AA.  
XX  
XX ADP29879;  
AC  
XX 01-DEC-2005 (revised)  
DT 12-AUG-2004 (first entry)  
XX  
XX Human secreted protein SEQ ID #1877.  
DE  
XX Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;  
KW cancer; inflammatory; immune; human secreted protein.  
XX  
XX Homo sapiens.  
OS  
XX WO2004035732-A2.  
PN  
XX 29-APR-2004.  
PD  
XX 28-AUG-2003; 2003WO-US026780.  
PF  
XX 29-AUG-2002; 2002US-0406576P.  
XX  
PR 29-AUG-2002; 2002US-0406579P.  
XX  
PR 29-AUG-2002; 2002US-0406585P.  
XX



CC recombinant expression cloning) techniques. The method involves providing  
 CC a nucleic acid library encoding genes/partial gene sequences of  
 CC pathogenic organisms, transforming/transfecting the library into host  
 CC cells, contacting the polypeptides expressed by the genes with autologous  
 CC antisera derived from an animal infected with, or has been infected with  
 CC the pathogen and purifying the nucleic acid encoding the polypeptide or  
 CC partial polypeptide binding to the antisera. Also included are the  
 CC nucleic acids and polypeptides isolated by the method, vectors and  
 CC transformed cells expressing them, a vaccine comprising the polypeptide  
 CC and the production of monoclonal antibodies against the polypeptides. The  
 CC protein and vaccine are useful for immunising an animal (preferably  
 CC human) against a pathogenic microbe. The proteins and antibodies are  
 CC useful for manufacturing a medicament for treating Staphylococcus aureus-  
 CC associated septicemia, food poisoning or skin disorders or  
 CC Staphylococcus epidermidis-associated septicemia, sepsis, meningitis,  
 CC endocarditis, tuberculosis, blood infections, sepsis, meningitis,  
 CC pneumonia, stomach ulcers, gonorrhea, necrotising fasciitis, impetigo,  
 CC Lyme's disease, gastro-enteritis, dysentery and shigellosis. The present  
 CC sequence is an *S. aureus* protein sequence associated with the antigenic  
 CC proteins of the invention. Note: The present sequence is included in the  
 CC sequence listing but is not mentioned anywhere else in the specification  
 XX  
 XX  
 SQ Sequence 541 AA;

Query Match 81.4%; Score 35; DB 5; Length 541;  
 Best Local Similarity 75.0%; Pred. No. 3.6e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8  
 Db :|||||  
 10 SQYEDYKQ 17

RESULT 8  
 AEA15502  
 ID AEA15502 standard; protein; 629 AA.  
 XX  
 AC AEA15502;  
 XX  
 DT 28-JUL-2005 (first entry)  
 XX  
 DE Human polypeptide #121.  
 XX  
 KW Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
 KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
 KW cytostatic; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005047534-A2.  
 XX  
 PD 26-MAY-2005.  
 XX  
 PF 15-OCT-2004; 2004WO-EP011599.  
 XX  
 PR 28-OCT-2003; 2003EP-00024565.  
 XX  
 PA (FARB ) BAYER HEALTHCARE AG.  
 XX  
 PI Wirtz R, Munnes M;  
 XX  
 DR WPI; 2005-372393/38.  
 DR N-PSDB; AEA15424.  
 XX  
 PT Predicting a response to cancer treatment by detecting at least 2  
 PT markers, which are genes or genomic nucleic acid sequences that are  
 PT located on one chromosomal region, which is altered in malignant  
 PT neoplasia.  
 XX  
 PS Claim 7; SEQ ID NO 464; 464pp; English.  
 XX  
 CC The invention relates to a method of predicting response to cancer  
 CC treatment comprising detection of at least 2 markers, where the markers

CC are genes and fragments or genomic nucleic acid sequences that are  
 CC located on one chromosomal region, which is altered in malignant  
 CC neoplasia. The invention also relates to a method for the prediction,  
 CC diagnosis or prognosis of malignant neoplasia, methods for detecting  
 CC deregulations in malignant neoplasia and breast cancer, a method of  
 CC determining the phenotype of a cell or tissue, a method for identifying  
 CC genomic regions which are altered on the chromosomal level and encode  
 CC genes that are linked by function and are differentially expressed in  
 CC malignant neoplasia and breast cancer, methods of screening for agents  
 CC which regulate the activity of a polypeptide or a polynucleotide and  
 CC antibodies that specifically bind to a full length or partial  
 CC polypeptide. The method is useful for predicting response to cancer  
 CC treatment. The methods and compositions are useful for predicting,  
 CC diagnosing, prognosing, preventing or treating malignant neoplasia  
 CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
 CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
 CC lung cancer. This sequence represents a human polypeptide used in the  
 CC scope of the invention.

SQ Sequence 629 AA;

Query Match 81.4%; Score 35; DB 9; Length 629;  
 Best Local Similarity 87.5%; Pred. No. 4.3e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8  
 Db :|||||  
 376 AQYEDIAQ 383

RESULT 9  
 AAU34342  
 ID AAU34342 standard; protein; 636 AA.

XX  
 AC AAU34342;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Staphylococcus aureus cellular proliferation protein #618.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 KW antibacterial; drug design.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US009180.  
 XX  
 PR 21-MAR-2000; 2000US-0191078P.  
 PR 23-MAY-2000; 2000US-0206848P.  
 PR 26-MAY-2000; 2000US-0207277P.  
 PR 23-OCT-2000; 2000US-0242578P.  
 PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 WPI; 2001-611495/70.  
 DR N-PSDB; AAS52201.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 PS Example 3; SEQ ID NO 5838; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence represents an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 636 AA;

Query Match 81.4%; Score 35; DB 4; Length 636;  
Best Local Similarity 75.0%; Pred. No. 4.4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQVEDYDQ 8  
:|||||  
Db 105 SQVEDYKQ 112

RESULT 10  
AAU37415  
ID AAU37415 standard; protein; 636 AA.

XX AC AAU37415;

XX DT 14-FEB-2002 (first entry)

XX DE Staphylococcus aureus cellular proliferation protein #1585.

XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
XX antibacterial; drug design.

XX OS Staphylococcus aureus.

XX FN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US009180.

XX PR 21-MAR-2000; 2000US-0191078P.

XX PR 23-MAY-2000; 2000US-0206848P.

XX PR 26-MAY-2000; 2000US-0207272P.

XX PR 23-OCT-2000; 2000US-0242578P.

XX PR 27-NOV-2000; 2000US-0253625P.

XX PR 22-DEC-2000; 2000US-0257931P.

XX PR 16-FEB-2001; 2001US-0269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX XX

XX PI Haeelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX FI Yamamoto RT, Xu HH;

XX XX WPI; 2001-611495/70.

XX DR N-PSDB; AAS55274.

XX XX

XX PT New polynucleotides for the identification and development of

CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence represents an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 636 AA;

Query Match 81.4%; Score 35; DB 4; Length 636;  
Best Local Similarity 75.0%; Pred. No. 4.4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQVEDYDQ 8  
:|||||  
Db 105 SQVEDYKQ 112

RESULT 10  
AAU37415  
ID AAU37415 standard; protein; 636 AA.

XX AC AAU37415;

XX DT 14-FEB-2002 (first entry)

XX DE Staphylococcus aureus cellular proliferation protein #1585.

XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
XX antibacterial; drug design.

XX OS Staphylococcus aureus.

XX FN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US009180.

XX PR 21-MAR-2000; 2000US-0191078P.

XX PR 23-MAY-2000; 2000US-0206848P.

XX PR 26-MAY-2000; 2000US-0207272P.

XX PR 23-OCT-2000; 2000US-0242578P.

XX PR 27-NOV-2000; 2000US-0253625P.

XX PR 22-DEC-2000; 2000US-0257931P.

XX PR 16-FEB-2001; 2001US-0269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX XX

XX PI Haeelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX FI Yamamoto RT, Xu HH;

XX XX WPI; 2001-611495/70.

XX DR N-PSDB; AAS55274.

XX XX

XX PT New polynucleotides for the identification and development of

CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence represents an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 636 AA;

Query Match 81.4%; Score 35; DB 4; Length 636;  
Best Local Similarity 75.0%; Pred. No. 4.4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQVEDYDQ 8  
:|||||  
Db 105 SQVEDYKQ 112

RESULT 11  
ADA89509  
ID ADA89509 standard; protein; 636 AA.

XX AC ADA89509;

XX DT 20-NOV-2003 (first entry)

XX DE Staphylococcus aureus antigenic protein #48.

XX KW antigenic protein; vaccine; Staphylococcus aureus; pathogenic organism;  
XX antibacterial; neuroprotective; immunosuppressive; antiinflammatory;  
KW antitumor; immunostimulant; ophthalmological; pathogenic microbe;  
KW bacteraemia; septic shock; organ infection; skin infection;  
KW bacterial basal colonisation; bacterial eye infection; septicaemia;  
KW tuberculosis; food poisoning; blood infection; peritonitis; endocarditis;  
KW sepsis; meningitis; pneumonia; stomach ulcer; gonorrhoea; toxic shock;  
KW necrotising fasciitis; impetigo; histoplasmosis; Lyme disease;  
KW gastro-enteritis; dysentery; shigellosis; skin disorder.

XX OS Staphylococcus aureus.

XX FN WO2003011899-A2.

XX PD 13-FEB-2003.

XX XX 02-AUG-2002; 2002WO-GB003606.

XX PF 02-AUG-2001; 2001GB-00018825.

XX PR 09-JAN-2002; 2002GB-00000349.

XX PA (UYSH-) UNIV SHEFFIELD.

XX FA (BIOS-) BIOSYNEXUS INC.

XX XX Foster S, Mond J, Clarke S, McDowell P, Brummel K;

XX DR WPI; 2003-256434/25.

XX XX

XX PT New antigenic polypeptides from Staphylococcus aureus or S. epidermidis;

XX PT useful as a vaccine for immunizing humans against e.g. bacteraemia, septic

XX PT shock, septicemia, tuberculosis, meningitis, pneumonia, gonorrhoea or

XX PT impetigo.

XX PS Claim 4; Page 133; 189pp; English.



CC The present invention describes an antigenic protein or its part, which  
 CC is for use as a vaccine. The antigenic protein is encoded by an isolated  
 CC DNA molecule of *Staphylococcus aureus* or *S. epidermidis* partial gene  
 CC sequences (designated dnaSA and dna SE, respectively; and which encodes a  
 CC protein expressed by a pathogenic organism. Also described: (1) a vaccine  
 CC composition comprising at least one antigenic protein; (2) a method of  
 CC immunising an animal against a disease or condition caused by a  
 CC pathogenic microbe by administering the antigenic protein or the vaccine;  
 CC (3) an antibody or its binding part obtainable by the method above; (4)  
 CC preparing a hybridoma cell line producing monoclonal antibodies; (5) a  
 CC hybridoma cell line produced by the method of (4); and (6) identifying  
 CC opsonic antigens expressed by a pathogenic microbe. The antigenic  
 CC proteins have antibacterial, neuroprotective, immunosuppressive,  
 CC antiinflammatory, antitumor, immunostimulant and ophthalmological  
 CC activities, and can be used in vaccines. The antigenic proteins or  
 CC vaccines can be used for immunising an animal (specifically a human)  
 CC against a disease or condition caused by a pathogenic microbe, e.g.  
 CC bacteraemia, septic shock, organ infection, skin infection, bacterial  
 CC basal colonisation, bacterial eye infections, septicaemia, tuberculosis,  
 CC bacteria-associated food poisoning, blood infections, peritonitis,  
 CC endocarditis, sepsis, meningitis, pneumonia, stomach ulcers, gonorrhoea,  
 CC strep throat, streptococcal-associated toxic shock, necrotising  
 CC fasciitis, impetigo, histoplasmosis, Lyme disease, gastro-enteritis,  
 CC dysentery, shigellosis, *S. aureus*-associated septicaemia, food-poisoning,  
 CC skin disorders, *S. epidermidis*-associated septicaemia, peritonitis or  
 CC endocarditis. The present sequence represents a *S. aureus* antigenic  
 CC protein sequence from the present invention.

XX Sequence 636 AA;

Query Match 81.4%; Score 35; DB 6; Length 636;

Best Local Similarity 75.0%; Pred. No. 4.4e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQYEDYQAQ 8

Db 105 SQYEDYKQ 112

RESULT 12

ABJ18985

ID ABJ18985 standard; protein; 636 AA.

XX AC ABJ18985;

XX DT 06-MAR-2003 (first entry)

XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 156.

XX KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;

XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;

XX KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;

XX KW autoimmune disease; HIV; hepatitis.

XX OS *Staphylococcus* sp.

XX PN WO200259148-A2.

XX PD 01-AUG-2002.

XX PF 21-JAN-2002; 2002WO-EP000546.

XX PR 26-JAN-2001; 2001AT-00000130.

XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

XX PI Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;

XX PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichart T, Hafner M;

XX PI Tempelmaier B;

XX DR WPI; 2003-075410/07.

XX PT Identifying, isolating and producing hyperimmune serum-reactive antigens

PT from a pathogen, for preparing vaccine or medicament for treating or  
 PT preventing e.g. staphylococcal infections, comprises providing antibody  
 PT preparation.

XX Example 7; Page 173; 252pp; English.

CC The invention relates to a novel method for identifying, isolating and  
 CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
 CC allergen, a tissue or host prone to auto-immunity, where the antigens are  
 CC used in a vaccine, comprises providing antibody preparation from a plasma  
 CC pool of a type of animal, or individual sera with antibodies against the  
 CC specific pathogen, tumour, allergen, tissue or host prone to auto-  
 CC immunity. The hyperimmune serum-reactive antigens comprising any of the  
 CC 62 sequences of 53-2261 amino acids fully defined in the specification,  
 CC or their hyperimmune fragments are useful for the manufacture of a  
 CC pharmaceutical preparation, particularly a vaccine against staphylococcal  
 CC infections or colonisation against *S. aureus* or *S. epidermidis*. The  
 CC preparation of antibodies is useful for the manufacture of a medicament  
 CC for treating or preventing staphylococcal infections or colonisation  
 CC against *S. aureus* or *S. epidermidis*. The antibody preparations may also  
 CC be used for diagnostic and imaging purposes. Other conditions that can be  
 CC treated include cancer, autoimmune diseases or infections caused by viral  
 CC (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This  
 CC sequence represents a staphylococcal protein relating to the method for  
 CC identifying and producing pathogen specific antigens of the invention

XX Sequence 636 AA;

Query Match 81.4%; Score 35; DB 6; Length 636;

Best Local Similarity 75.0%; Pred. No. 4.4e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AQYEDYQAQ 8

Db 105 SQYEDYKQ 112

RESULT 13

ABM72027

ID ABM72027 standard; protein; 636 AA.

XX AC ABM72027;

XX DT 20-NOV-2003 (first entry)

XX DE *Staphylococcus aureus* protein #1267.

XX KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;

XX KW enzymatic assay; antibiotic target.

XX OS *Staphylococcus aureus*.

XX PN WO200294868-A2.

XX PD 28-NOV-2002.

XX PF 27-MAR-2002; 2002WO-IB002637.

XX PR 27-MAR-2001; 2001GB-00007661.

XX PA (CHIR-) CHIRON SPA.

XX PI Massignani V, Mora M, Scarselli M;

XX DR WPI; 2003-120786/11.

XX XX N-PSDB; ACF73587.

XX PT New *Staphylococcus aureus* protein, useful as a vaccine for treating or  
 PT preventing *Staphylococcal* infection, specifically an infection caused by  
 PT *S. aureus*, e.g. sepsis.

XX PS Claim 1; SEQ ID NO 2534; 49pp; English.

The invention relates to novel genes and encoded proteins from *Staphylococcus aureus*. A composition comprising the *S. aureus* protein, a nucleic acid encoding the protein, or an antibody to the protein, is useful as a pharmaceutical, particularly as a vaccine for treating or preventing infection due to *Staphylococcus* bacteria, specifically an infection caused by *S. aureus*. The composition is particularly useful for treating or preventing sepsis in a patient. The composition can also be used for diagnostics. The protein is also used in an assay for enzymatic studies and as a target for antibiotics. This sequence represents one of the novel *S. aureus* proteins of the invention.

polypeptide. The method is useful for predicting response to cancer treatment. The methods and compositions are useful for predicting, diagnosing, prognosing, preventing or treating malignant neoplasia including breast cancer, ovarian cancer, gastric cancer, colon cancer, esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell lung cancer. This sequence represents a human polypeptide used in the scope of the invention.

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AQYEDYAO 8  
| | | | |  
Db 105 AQYEDKAO 112

Search completed: July 5, 2006, 18:43:49  
Job time : 71.5146 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 58.5631 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-10  
Perfect score: 43  
Sequence: 1 AQYEDYQAQ 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	43	100.0	8	5	US-10-774-928-10
2	38	88.4	98	4	US-10-437-963-128532
3	35	81.4	9	5	US-10-774-928-2
4	35	81.4	541	4	US-10-311-879-30
5	35	81.4	636	3	US-09-815-242-5838
6	35	81.4	636	3	US-09-815-242-13008
7	35	81.4	636	5	US-10-470-048B-156
8	35	81.4	636	5	US-10-485-517-170
9	33	76.7	3241	4	US-10-282-122A-49924
10	32	74.4	52	4	US-10-424-599-190535
11	32	74.4	143	4	US-10-425-115-362330
12	32	74.4	176	6	US-11-097-143-28200
13	32	74.4	201	4	US-10-425-115-361451
14	32	74.4	417	4	US-10-437-963-167445
15	32	74.4	587	3	US-09-815-242-5694
16	32	74.4	603	3	US-09-815-242-12667
17	32	74.4	603	4	US-10-282-122A-44316
18	32	74.4	1132	4	US-10-437-963-169428
19	31	72.1	11	3	US-09-946-678-7
20	31	72.1	43	4	US-10-437-963-121052
21	31	72.1	82	3	US-09-864-761-41119
22	31	72.1	106	5	US-10-220-335-631
23	31	72.1	131	5	US-10-793-626-1872
24	31	72.1	143	4	US-10-425-115-345358
25	31	72.1	150	4	US-10-724-972A-5799
26	31	72.1	179	6	US-11-097-143-15546
27	31	72.1	200	4	US-10-424-599-184988

28	31	72.1	202	6	US-11-097-143-26418	Sequence 26418, A
29	31	72.1	279	5	US-10-467-657-3032	Sequence 3032, Ap
30	31	72.1	280	3	US-09-815-242-12479	Sequence 12479, A
31	31	72.1	290	4	US-10-767-701-45172	Sequence 45172, A
32	31	72.1	290	4	US-10-425-115-303461	Sequence 303461, A
33	31	72.1	300	4	US-10-437-963-150328	Sequence 150328, A
34	31	72.1	312	6	US-11-188-298-18557	Sequence 18557, A
35	31	72.1	327	4	US-10-408-765A-1475	Sequence 1475, Ap
36	31	72.1	341	4	US-10-425-114-46394	Sequence 46394, A
37	31	72.1	433	6	US-11-188-298-2247	Sequence 2247, Ap
38	31	72.1	442	5	US-10-467-657-3900	Sequence 3900, Ap
39	31	72.1	444	4	US-10-282-122A-72394	Sequence 72394, A
40	31	72.1	448	6	US-11-188-298-16905	Sequence 16905, A
41	31	72.1	481	6	US-11-188-298-5542	Sequence 5542, Ap
42	31	72.1	484	6	US-11-188-298-16281	Sequence 16281, A
43	31	72.1	490	6	US-11-188-298-17034	Sequence 17034, A
44	31	72.1	504	3	US-09-765-205-40	Sequence 40, Appl
45	31	72.1	504	5	US-10-347-669-40	Sequence 40, Appl

## ALIGNMENTS

## RESULT 1

US-10-774-928-10  
; Sequence 10, Application US/10774928  
; Publication NO. US20040241168A1  
; GENERAL INFORMATION:

APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Psoriasis

FILE REFERENCE: 302178

CURRENT APPLICATION NUMBER: US/10/774,928

CURRENT FILING DATE: 2004-02-09

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn version 3.0

SEQ ID NO 10

LENGTH: 8

TYPE: PRT

ORGANISM: Leishmania

US-10-774-928-10

Query Match 100.0%; Score 43; DB 5; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.9e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8

DB 1 AQYEDYQAQ 8

## RESULT 2

US-10-437-963-128532

; Sequence 128532, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 128532

LENGTH: 98

TYPE: PRT

```
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_30877C.1.pep
US-10-437-963-128532

Query Match      88.4%; Score 38; DB 4; Length 98;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYQAQ 8
Db      36 ANYEDYQAQ 43

RESULT 3
US-10-774-928-2
; Sequence 2, Application US/10774928
; Publication No. US20040241168A1
; GENERAL INFORMATION:
; APPLICANT: O'Daly, Jose
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission
; FILE REFERENCE: Psoriasis
; CURRENT APPLICATION NUMBER: 302178
; CURRENT FILING DATE: 2004-02-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Leishmania
US-10-774-928-2

Query Match      81.4%; Score 35; DB 5; Length 9;
Best Local Similarity 87.5%; Pred. No. 1.9e+06;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYQAQ 8
Db      1 AQYEDYQAQ 8

RESULT 4
US-10-311-879-30
; Sequence 30, Application US/103111879
; Publication No. US20030186275A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; TITLE OF INVENTION: Antigenic Peptides
; FILE REFERENCE: toxin
; CURRENT APPLICATION NUMBER: US/10/311,879
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-311-879-30

Query Match      81.4%; Score 35; DB 4; Length 541;
Best Local Similarity 75.0%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYQAQ 8
Db      10 SQYEDYKQ 17

RESULT 5
US-09-815-242-5838
; Sequence 5838, Application US/09815242
; Patent No. US20020061569A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5838
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5838

Query Match      81.4%; Score 35; DB 3; Length 636;
Best Local Similarity 75.0%; Pred. No. 5.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYQAQ 8
Db      105 SQYEDYKQ 112

RESULT 6
US-09-815-242-13008
; Sequence 13008, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: PastSeq for Windows Version 4.0  
;; SEQ ID NO 13008  
;; LENGTH: 636  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-13008

Query Match 81.4%; Score 35; DB 3; Length 636;  
Best Local Similarity 75.0%; Pred. No. 5.9e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8  
:|||||  
Db 105 SQYEDYKQ 112

## RESULT 7

US-10-470-048B-156  
;; Sequence 156, Application US/10470048B  
;; Publication No. US2005003744A1  
;; GENERAL INFORMATION:  
;; APPLICANT: MEINKE ET AL.  
;; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF  
;; FILE REFERENCE: SOINN-035US  
;; CURRENT APPLICATION NUMBER: US/10/470,048B  
;; CURRENT FILING DATE: 2003-07-25  
;; NUMBER OF SEQ ID NOS: 603  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 156  
;; LENGTH: 636  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus aureus  
US-10-470-048B-156

Query Match 81.4%; Score 35; DB 5; Length 636;  
Best Local Similarity 75.0%; Pred. No. 5.9e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8  
:|||||  
Db 105 SQYEDYKQ 112

## RESULT 8

US-10-485-517-170  
;; Sequence 170, Application US/10485517  
;; Publication No. US2005025629A1  
;; GENERAL INFORMATION:  
;; APPLICANT: University of Sheffield  
;; APPLICANT: Biosynex Incorporated  
;; APPLICANT: Foster, Simon  
;; APPLICANT: Mond, James  
;; TITLE OF INVENTION: Antigenic Polypeptides  
;; FILE REFERENCE: P100629WO  
;; CURRENT APPLICATION NUMBER: US/10/485,517  
;; CURRENT FILING DATE: 2004-02-02  
;; PRIOR APPLICATION NUMBER: GB 0118825.9  
;; PRIOR FILING DATE: 2001-08-02  
;; PRIOR APPLICATION NUMBER: GB 0200349.9  
;; PRIOR FILING DATE: 2002-01-09  
;; NUMBER OF SEQ ID NOS: 424  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 170  
;; LENGTH: 636  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus aureus  
US-10-485-517-170

Query Match 81.4%; Score 35; DB 5; Length 636;  
Best Local Similarity 75.0%; Pred. No. 5.9e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8  
:|||||  
Db 105 SQYEDYKQ 112

## RESULT 9

US-10-282-122A-49924  
;; Sequence 49924, Application US/10282122A  
;; Publication No. US20040029129A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, Liangsu  
;; APPLICANT: Zamudio, Carlos  
;; APPLICANT: Malone, Cheryl  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari  
;; APPLICANT: Zyskind, Judith  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John  
;; APPLICANT: Carr, Grant  
;; APPLICANT: Yamamoto, Robert  
;; APPLICANT: Forsyth, R.  
;; APPLICANT: Xu, H.  
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
;; FILE REFERENCE: ELITRA.034A  
;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 49924  
;; LENGTH: 3241  
;; TYPE: PRT  
;; ORGANISM: Burkholderia fungorum  
US-10-282-122A-49924

Query Match 76.7%; Score 33; DB 4; Length 3241;  
Best Local Similarity 85.7%; Pred. No. 6.7e+03;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QYEDYQAQ 8  
:|||||  
Db 795 QYADYQAQ 801

## RESULT 10

US-10-424-599-190535  
;; Sequence 190535, Application US/10424599  
;; Publication No. US20040031072A1

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 190535
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_143072C.1.pep
US-10-424-599-190535

Query Match          74.4%; Score 32; DB 4; Length 52;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QYEDYQAQ 8
Db      40 QYEHYQAQ 46

RESULT 11
US-10-425-115-362330
; Sequence 362330, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 362330
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_93620C.1.pep
US-10-425-115-362330

Query Match          74.4%; Score 32; DB 4; Length 143;
Best Local Similarity 62.5%; Pred. No. 4.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYQAQ 8
Db      46 SQYKDYVQ 53

RESULT 12
US-11-097-143-28200
; Sequence 28200, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
```

```
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28200
; LENGTH: 176
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-28200

Query Match          74.4%; Score 32; DB 6; Length 176;
Best Local Similarity 62.5%; Pred. No. 5.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYQAQ 8
Db      162 AQYDDYVK 169

RESULT 13
US-10-425-115-361451
; Sequence 361451, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 361451
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92821C.1.pep
US-10-425-115-361451

Query Match          74.4%; Score 32; DB 4; Length 201;
Best Local Similarity 62.5%; Pred. No. 6.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYQAQ 8
Db      107 SQFEDYVQ 114

RESULT 14
US-10-437-963-167445
; Sequence 167445, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```



APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 167445  
LENGTH: 417  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_66056C.1.pap  
US-10-437-963-167445

Query Match 74.4%; Score 32; DB 4; Length 417;  
Best Local Similarity 85.7%; Pred. No. 1.3e+03;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QYEDYQAQ 8  
DB 211 QYETYQAQ 217

RESULT 15  
US-09-815-242-5694  
Sequence 5694, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5694  
LENGTH: 587  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5694

Query Match 74.4%; Score 32; DB 3; Length 587;  
Best Local Similarity 62.5%; Pred. No. 1.9e+03;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQYDYQAQ 8

Db 478 SQKDYAK 485  
:|:|:|:

Search completed: July 5, 2006, 20:51:58  
Job time : 59.5631 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 4.58252 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-10  
Perfect score: 43  
Sequence: 1 AQYEDYQAQ 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New.\*  
1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	35	81.4	541	7 US-11-256-173-30
2	35	81.4	636	6 US-10-471-571A-2534
3	32	74.4	143	6 US-10-449-902-53316
4	32	74.4	417	6 US-10-449-902-51027
5	32	74.4	603	6 US-10-471-571A-3718
6	31	72.1	300	6 US-10-449-902-46061
7	31	72.1	475	7 US-11-174-307B-160
8	31	72.1	548	6 US-10-449-902-29496
9	31	72.1	585	6 US-10-449-902-53167
10	31	72.1	852	6 US-10-449-902-51138
11	31	72.1	1127	6 US-10-449-902-53809
12	30	69.8	368	6 US-10-449-902-31471
13	30	69.8	368	6 US-10-449-902-54507
14	30	69.8	483	7 US-11-105-233-180
15	30	69.8	797	6 US-10-449-902-41261
16	29	67.4	386	7 US-11-256-173-16
17	29	67.4	389	7 US-11-264-558-5
18	29	67.4	467	7 US-11-264-558-4
19	29	67.4	593	7 US-11-293-697-4773
20	29	67.4	619	6 US-10-471-571A-3914
21	29	67.4	944	6 US-10-449-902-47242
22	29	67.4	2397	6 US-10-471-571A-2632
23	28	65.1	103	6 US-10-953-349-26513
24	28	65.1	110	7 US-11-342-171-4
25	28	65.1	125	6 US-10-530-879-10

26	28	65.1	217	6	US-10-953-349-35566	Sequence 35566, A
27	28	65.1	222	6	US-10-953-349-12891	Sequence 12891, A
28	28	65.1	245	6	US-10-953-349-35565	Sequence 35565, A
29	28	65.1	258	6	US-10-953-349-18067	Sequence 18067, A
30	28	65.1	289	6	US-10-953-349-12890	Sequence 12890, A
31	28	65.1	309	6	US-10-449-902-31890	Sequence 31890, A
32	28	65.1	309	6	US-10-449-902-49974	Sequence 49974, A
33	28	65.1	311	6	US-10-953-349-1338	Sequence 1338, Ap
34	28	65.1	312	6	US-10-953-349-1337	Sequence 1337, Ap
35	28	65.1	314	6	US-10-953-349-1336	Sequence 1336, Ap
36	28	65.1	320	6	US-10-449-902-52190	Sequence 52190, A
37	28	65.1	328	6	US-10-449-902-31205	Sequence 31205, A
38	28	65.1	349	6	US-10-953-349-12889	Sequence 12889, A
39	28	65.1	371	6	US-10-530-879-3	Sequence 3, Appli
40	28	65.1	372	6	US-10-953-349-11032	Sequence 11032, A
41	28	65.1	452	6	US-10-953-349-11031	Sequence 11031, A
42	28	65.1	453	6	US-10-953-349-11030	Sequence 11030, A
43	28	65.1	535	7	US-11-293-697-4336	Sequence 4336, Ap
44	28	65.1	623	7	US-11-342-171-64	Sequence 64, Appl
45	28	65.1	971	6	US-10-449-902-53062	Sequence 53062, A

## ALIGNMENTS

RESULT 1  
US-11-256-173-30  
; Sequence 30, Application US/11256173  
; Publication No. US20060140979A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; TITLE OF INVENTION: Antigenic Peptides  
; FILE REFERENCE: toxin  
; CURRENT APPLICATION NUMBER: US/11/256,173  
; CURRENT FILING DATE: 2005-10-24  
; PRIOR APPLICATION NUMBER: US/10/311,879  
; PRIOR FILING DATE: 2003-03-18  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 30  
; LENGTH: 541  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-11-256-173-30

Query Match 81.4%; Score 35; DB 7; Length 541;  
Best Local Similarity 75.0%; Pred. No. 13;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQYEDYQAQ 8  
DB 10 SQYEDYKQ 17

RESULT 2  
US-10-471-571A-2534  
; Sequence 2534, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON Spa  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927WO  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 2534  
; LENGTH: 636  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
; FEATURE:

```
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(636)
; OTHER INFORMATION: staphylocoagulase precursor
US-10-471-571A-2534

Query Match      81.4%; Score 35; DB 6; Length 636;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDYQAQ 8
Db 105 SQYEDYKQ 112

RESULT 3
US-10-449-902-53316
; Sequence 53316, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53316
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53316

Query Match      74.4%; Score 32; DB 6; Length 143;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QYEDYQAQ 8
Db 127 EYEEYQAQ 133

RESULT 4
US-10-449-902-51027
; Sequence 51027, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51027
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-51027
```

```
Query Match      74.4%; Score 32; DB 6; Length 417;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QYEDYQAQ 8
Db 211 QYETVQAQ 217

RESULT 5
US-10-471-571A-3718
; Sequence 3718, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 3718
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(603)
; OTHER INFORMATION: hypothetical protein
US-10-471-571A-3718

Query Match      74.4%; Score 32; DB 6; Length 603;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQYEDYQAQ 8
Db 488 SQYKDYAK 495

RESULT 6
US-10-449-902-46061
; Sequence 46061, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46061
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-46061

Query Match      72.1%; Score 31; DB 6; Length 300;
Best Local Similarity 62.5%; Pred. No. 42;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQYEDYQAQ 8
Db 111 : : : :
```

```
Db      104 AQFEDYKE 111

RESULT 7
US-11-174-307B-160
; Sequence 160, Application US/11174307B
; Publication No. US20060143729A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY
; TITLE OF INVENTION: USEFUL FOR MODIFYING PLANT CHARACTERISTICS
; FILE REFERENCE: 2750-1601PUS2
; CURRENT APPLICATION NUMBER: US/11/174, 307B
; CURRENT FILING DATE: 2003-06-30
; PRIOR APPLICATION NUMBER: 60/583,671
; PRIOR FILING DATE: 2004-06-30
; PRIOR APPLICATION NUMBER: 60/583,781
; PRIOR FILING DATE: 2004-06-30
; PRIOR APPLICATION NUMBER: 60/583,651
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 5544
; SEQ ID NO 160
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: Pfam Name: AAA; Pfam Description: ATPase family associated
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 10176992; NR Description: contains similarity to
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 9294271; NR Description: mitochondrial
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 9294274; NR Description: mitochondrial
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 61656802; NR Description: Cell Division Protein
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 45735466; NR Description: putative ATPase
US-11-174-307B-160

Query Match      72.1%; Score 31; DB 7; Length 475;
Best Local Similarity 83.3%; Pred. No. 68;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AQYEDY 6
       :|||||
Db      65 SQYEDY 70

RESULT 8
US-10-449-902-29496
; Sequence 29496, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A02051-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29496
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-29496

Query Match      72.1%; Score 31; DB 6; Length 548;
Best Local Similarity 71.4%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 QYEDYAQ 8
       ||:|||
Db      244 QYQDYQQ 250

RESULT 9
US-10-449-902-53167
; Sequence 53167, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A02051-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53167
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53167

Query Match      72.1%; Score 31; DB 6; Length 585;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 QYEDYAQ 8
       ||:|||
Db      281 QYQDYQQ 287

RESULT 10
US-10-449-902-51138
; Sequence 51138, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A02051-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
```

```
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51138
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-51138

Query Match      72.1%; Score 31; DB 6; Length 852;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QYEDYQAQ 8
Db      626 QYQDYQQ 632

RESULT 11
US-10-449-902-53809
; Sequence 53809, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53809
; LENGTH: 1127
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53809

Query Match      72.1%; Score 31; DB 6; Length 1127;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QYEDYQAQ 8
Db      823 QYQDYQQ 829

RESULT 12
US-10-449-902-31471
; Sequence 31471, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31471
; LENGTH: 368
; TYPE: PRT
```

```
; ORGANISM: Oryza sativa
US-10-449-902-31471

Query Match      69.8%; Score 30; DB 6; Length 368;
Best Local Similarity 71.4%; Pred. No. 81;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AQYEDYA 7
Db      62 SQWEDYA 68

RESULT 13
US-10-449-902-54507
; Sequence 54507, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54507
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-54507

Query Match      69.8%; Score 30; DB 6; Length 368;
Best Local Similarity 71.4%; Pred. No. 81;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AQYEDYA 7
Db      62 SQWEDYA 68

RESULT 14
US-11-105-233-180
; Sequence 180, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; FILE REFERENCE: Tumors
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-180

Query Match      69.8%; Score 30; DB 7; Length 483;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AQYEDYA 7
Db      265 AQYEDIA 271
```

RESULT 15  
US-10-449-902-41261  
; Sequence 41261, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 41261  
; LENGTH: 797  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-41261

Query Match 69.8%; Score 30; DB 6; Length 797;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QYEDY 6  
DB 491 QYEDY 495

Search completed: July 5, 2006, 20:53:01  
Job time : 4.58252 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 9.70874 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928a-11  
Perfect score: 44  
Sequence: 1 YEDEINN 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	84.1	484	2 E64432	spore coat polys
2	37	84.1	506	2 S37156	NADPH-ferrihemopro
3	37	84.1	683	2 T10720	NADPH-ferrihemopro
4	37	84.1	692	2 T05582	NADPH-ferrihemopro
5	37	84.1	692	2 S37159	NADPH-ferrihemopro
6	36	81.8	403	2 T28416	ORF MSV255 leucine
7	36	81.8	597	2 S43865	phosphoglucosylase
8	35	79.5	310	2 S43865	cytokeratin 8, typ
9	35	79.5	384	2 I61769	keratin 6d, type I
10	35	79.5	390	2 A60093	cytokeratin, type I
11	35	79.5	419	1 KRXL2B	keratin, 64k type
12	35	79.5	425	1 KRXL2A	keratin, 64k type
13	35	79.5	461	2 A43782	keratin, type II -
14	35	79.5	469	2 S05602	keratin K7, type I
15	35	79.5	483	2 A34720	keratin 8, type II
16	35	79.5	487	2 JT0407	keratin 8, type II
17	35	79.5	489	2 B24177	keratin, 55k type
18	35	79.5	490	2 JS0658	cytokeratin EndoA
19	35	79.5	502	2 A23547	keratin, type II c
20	35	79.5	513	2 S08381	keratin, 58k type
21	35	79.5	520	2 JS0291	intermediate filam
22	35	79.5	553	2 I59009	epidermal keratin
23	35	79.5	564	1 KRHUEA	keratin 6a, type I
24	35	79.5	564	1 KRHUEB	keratin 6b, type I
25	35	79.5	564	2 I61770	keratin 6c, type I
26	35	79.5	564	2 I61768	keratin 6c, type I
27	35	79.5	564	2 I61771	keratin 6f, type I
28	35	79.5	581	1 KRMS2	keratin, type II c
29	35	79.5	590	2 A29904	keratin 5, type II

30	35	79.5	629	2 A29666	keratin, 65K type
31	35	79.5	629	2 S42629	keratin K3 - rabbi
32	35	79.5	638	2 I53169	cytokeratin 2 - hu
33	35	79.5	643	1 KRHU2	keratin 1, type II
34	35	79.5	645	2 A44861	keratin, 67K type
35	35	79.5	1029	2 F86359	hypothetical prote
36	35	79.5	1161	2 T18400	glutamate-cysteine
37	34	77.3	223	2 T00150	hypothetical prote
38	34	77.3	453	2 S31313	hypothetical prote
39	34	77.3	1254	2 S46636	hypothetical prote
40	34	77.3	2183	2 T37218	hypothetical prote
41	33	75.0	146	2 D86419	hypothetical prote
42	33	75.0	236	2 H71056	probable arylaloni
43	33	75.0	240	2 H83970	hypothetical prote
44	33	75.0	410	2 S77585	pilus assembly pro
45	33	75.0	410	2 H81210	pilus assembly pro

ALIGNMENTS

RESULT 1

E64432  
spore coat polysaccharide biosynthesis protein G homolog - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: E64432  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, A.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, H.O.; Woeese, C. Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woeese, C. A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: E64432  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-484 <BUL>  
C:Cross-references: UNIPROT:Q58462; UNIPARC:UPI0000139F8D; GB:U67549; GB:L77117; NID:9282; C:Genetics:  
A:Map position: REV1004535-1003081

Query Match 84.1%; Score 37; DB 2; Length 484;  
Best Local Similarity 85.7%; Pred. No. 18;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
|||:|  
DB 211 YEDELNN 217

RESULT 2

S37156  
NADPH-ferrihemoprotein reductase (EC 1.6.2.4) - Jerusalem artichoke (fragment)  
C:Species: Helianthus tuberosus (Jerusalem artichoke)  
C>Date: 09-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S37156  
R:Hasenfratz, M.; Jeltsch, J.; Benveniste, I.; Lesot, A.; Durst, F. submitted to the EMBL Data Library, September 1993  
A:Description: Multiple forms of NADPH-cytochrome P450 reductase in higher plants.  
A:Reference number: S37156  
A:Accession: S37156  
A:Molecule type: mRNA  
A:Residues: 1-506 <HAS>  
C:Cross-references: UNIPROT:Q96560; UNIPARC:UPI000009D6B1; EMBL:Z26251; NID:g1359893; PFI C:Superfamily: NADPH-ferrihemoprotein reductase; flavodoxin homology; NADPH-ferrihemoprotein C:Keywords: flavoprotein; NADP; oxidoreductase  
F;1-506/Domain: NADPH-ferrihemoprotein reductase homology (fragment) <FEH>  
F;1-52/Domain: Flavodoxin homology (fragment) <FLX>

Query Match 84.1%; Score 37; DB 2; Length 506;  
Best Local Similarity 85.7%; Pred. No. 19;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDELNN 7  
| | | | |  
Db 405 YEDELNN 411

## RESULT 3

T10720  
NADPH-ferrihemoprotein reductase (EC 1.6.2.4) - opium poppy  
N:Alternate names: ferrihemoprotein P-450 reductase; NADP-cytochrome reductase  
C:Species: Papaver somniferum (Opium poppy)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T10720  
R:Rosco, A.; Pauli, H.H.; Priesner, W.; Kutchan, T.M.  
Arch. Biochem. Biophys. 348, 369-377, 1997  
A:Title: Cloning and heterologous expression of NADPH-cytochrome P450 reductases from the  
A:Reference number: Z17096; MUID:98096363; PMID:9434750  
A:Accession: T10720  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-683 <ROS>  
A:Cross-references: UNIPROT:O24424; UNIPARC:UPI000009D902; EMBL:U67185; NID:G2580496; PDB:1W01  
C:Function:  
C:Superfamily: NADPH-ferrihemoprotein reductase; flavodoxin homology; NADPH-ferrihemoprotein reductase  
C:Keywords: flavoprotein; NADP; oxidoreductase  
F:80-682/Domain: NADPH-ferrihemoprotein reductase homology <FEH>  
F:82-230/Domain: flavodoxin homology <FLX>

Query Match 84.1%; Score 37; DB 2; Length 683;  
Best Local Similarity 85.7%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDELNN 7  
| | | | |  
Db 582 YEDELNN 588

## RESULT 4

T05582  
NADPH-ferrihemoprotein reductase (EC 1.6.2.4) ATR1 - Arabidopsis thaliana  
N:Alternate names: protein F22K18.280  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T05582; S21530  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohenseil, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05582  
A:Molecule type: DNA  
A:Residues: 1-692 <BEV>  
A:Cross-references: UNIPROT:Q9SB48; UNIPARC:UPI00000A72AC; EMBL:AL035356  
A:Experimental source: cultivar Columbia; BAC clone F22K18  
R:Mignote-Vieux, C.; Kazmaier, M.; Lacroute, F.; Pompon, D.M.  
submitted to the EMBL Data Library, May 1992  
A:Reference number: S21530  
A:Accession: S21530

Query Match 84.1%; Score 37; DB 2; Length 692;  
Best Local Similarity 85.7%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## A:Gene: ATR1

A:Map position: 4  
A:Introns: 101/3; 121/3; 147/2; 166/3; 194/3; 207/1; 230/2; 294/2; 325/2; 400/3; 428/3;  
A:Note: F22K18.280  
C:Superfamily: NADPH-ferrihemoprotein reductase; flavodoxin homology; NADPH-ferrihemoprotein reductase  
C:Keywords: flavoprotein; NADP; oxidoreductase  
F:85-691/Domain: NADPH-ferrihemoprotein reductase homology <FEH>  
F:87-235/Domain: flavodoxin homology <FLX>

QY 1 YEDELNN 7  
| | | | |  
Db 591 YEDELNN 597

## RESULT 5

S37159  
NADPH-ferrihemoprotein reductase (EC 1.6.2.4) - spring vetch  
C:Species: Vicia sativa (spring vetch, tare)  
C:Date: 20-May-1994 #sequence\_revision 03-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: S37159  
R:Benveniste, I.; Beque-Kirn, C.; Lesot, A.; Hasenfratz, M.; Durst, F.  
submitted to the EMBL Data Library, September 1993  
A:Description: Isolation and characterization of a cDNA encoding an NADPH-cytochrome P450  
A:Reference number: S37159  
A:Accession: S37159  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-692 <BEN>  
A:Cross-references: UNIPROT:Q43235; UNIPARC:UPI00000A4241; EMBL:Z26252; NID:G400531; PDB:1W01  
C:Superfamily: NADPH-ferrihemoprotein reductase; flavodoxin homology; NADPH-ferrihemoprotein reductase  
C:Keywords: flavoprotein; NADP; oxidoreductase  
F:83-691/Domain: NADPH-ferrihemoprotein reductase homology <FEH>  
F:85-234/Domain: flavodoxin homology <FLX>

Query Match 84.1%; Score 37; DB 2; Length 692;  
Best Local Similarity 85.7%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDELNN 7  
| | | | |  
Db 591 YEDELNN 597

## RESULT 6

T28416  
ORF MSV255 leucine rich repeat gene family protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T28416  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.

A:Reference number: Z20484; MUID:95102612; PMID:9847359

A:Accession: T28416

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-403 <AFO>

A:Cross-references: UNIPROT:Q9YV17; UNIPARC:UPI00000F4752; EMBL:AF063866; NID:G4049647; 1

C:Genetics:

A:Note: MSV255

Query Match 81.8%; Score 36; DB 2; Length 403;  
Best Local Similarity 87.5%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDELNNK 8  
| | | | |  
Db 10 YEDIINN 17

## RESULT 7

D70100  
phosphoglucosyltransferase (femD) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
C:Accession: D70100  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, S.; Peterson, J.; Kraljic, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A;Reference number: A70100; MUID:98065943; PMID:9403685  
A;Accession: D70100  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-597 <KLE>  
A;Cross-references: UNIPROT:O51037; UNIPARC:UPI00000572F1; GB:AE001115; GB:AE000783; NID  
A;Experimental source: strain B31

Query Match 81.0%; Score 36; DB 2; Length 597;  
Best Local Similarity 75.0%; Pred. No. 36;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINNK 8

DB 216 YENDINNK 223

#### RESULT 8

A;Title: cyokeratin 8, type II - potoro (fragment)  
A;Species: Potorous tridactylus (potoroo, long-nosed rat kangaroo)  
C;Date: 20-Feb-1995 #sequence\_revision 06-Dec-1996 #text\_change 09-Jul-2004  
C;Accession: S43865; S40172  
R;Boettger, V.; Lane, E.B.  
J. Mol. Biol. 235, 61-67, 1994

A;Title: A monoclonal antibody epitope on keratin 8 identified using a phage peptide lib  
A;Reference number: S43865; MUID:94118290; PMID:7507169

A;Accession: S43865

A;Molecule type: mRNA

A;Residues: 1-310 <BOE>

A;Cross-references: UNIPROT:Q28810; UNIPARC:UPI0000086BF0; EMBL:X70987; NID:g437902; PID  
A;Experimental source: kidney epithelium cells  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 79.5%; Score 35; DB 2; Length 310;

Best Local Similarity 75.0%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8

DB 180 YEDEINNK 187

#### RESULT 9

A;Title: keratin 6d, type II - human (fragment)  
A;Species: Homo sapiens (man)  
C;Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
C;Accession: I61769  
R;Takahashi, K.; Paladini, R.D.; Coulombe, P.A.  
J. Biol. Chem. 270, 18581-18592, 1995

A;Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A;Reference number: A57398; MUID:95355491; PMID:7543104

A;Accession: I61769

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-384 <RES>

A;Cross-references: UNIPROT:P48667; UNIPARC:UPI000012DB17; GB:L42602; NID:g914823; GB:L4

28; GB:L42608; NID:g914829; GB:L42609; NID:g914830; GB:L42610; NID:g1488252; PIDN:AAB606

C;Genetics:

A;Gene: KRT6D

A;Introns: 72/2; 92/3; 124/3; 179/3; 221/3; 295/2; 307/1

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil

Query Match 79.5%; Score 35; DB 2; Length 384;

Best Local Similarity 75.0%; Pred. No. 36;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8

DB 180 YEDEINNK 187

Db 73 YEDEINNK 80

#### RESULT 10

A60093

A;Title: cyokeratin, type II, early ectodermal - chicken (fragment)

C;Species: Gallus gallus (chicken)

C;Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 30-Sep-1993

C;Accession: A60093

R;Charlebois, T.S.; Spencer, D.H.; Tarkington, S.K.; Henry, J.J.; Grainger, R.M.

A;Title: Isolation of a chick cyokeratin cDNA clone indicative of regional specialization

A;Reference number: A60093; MUID:90276238; PMID:1693557

A;Accession: A60093

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-390 <CHA>

A;Cross-references: UNIPARC:UPI00001774C6

C;Superfamily: cytoskeletal keratin

C;Keywords: intermediate filament

Query Match 79.5%; Score 35; DB 2; Length 390;

Best Local Similarity 75.0%; Pred. No. 36;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8

DB 100 YEDEINNK 107

#### RESULT 11

KRXL2B

A;Title: keratin, 64K type II cytoskeletal (clone pUF164) - African clawed frog (fragment)

N;Alternate names: 64-kDa type II keratin

C;Species: Xenopus laevis (African clawed frog)

C;Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004

C;Accession: A02952

R;Hoffmann, W.; Franz, J.K.; Franke, W.W.

J. Mol. Biol. 184, 713-724, 1985

A;Title: Amino acid sequence microheterogeneities of basic (type II) cyokeratins of Xen

A;Reference number: A92914; MUID:86011576; PMID:2413219

A;Accession: A02952

A;Molecule type: mRNA

A;Residues: 1-419 <HOF>

A;Cross-references: UNIPROT:P04265; UNIPARC:UPI000012DB07; GB:X02895; GB:M13954; NID:g64f

A;Experimental source: clone pUF164

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil; intermediate filament

F;1-275/Domain: rod (fragment) <ROD>

F;1-9/Region: linker 1

F;10-110/Region: coil 1B

F;11-127/Region: linker 12

F;128-146/Region: coil 2A

F;147-154/Region: linker 2

F;155-275/Region: coil 2B

F;213/Region: stutter

F;276-419/Domain: tail <END>

F;276-399/Region: H2 and V2 subdomains

F;400-419/Region: E2 subdomain

Query Match 79.5%; Score 35; DB 1; Length 419;

Best Local Similarity 75.0%; Pred. No. 39;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8

DB 52 YEDEINNK 59

#### RESULT 12

KRXL2A

A;Title: keratin, 64K type II cytoskeletal (clone pUF23) - African clawed frog (fragment)

N;Alternate names: 64-kDa type II keratin

C:Species: Xenopus laevis (African clawed frog)  
C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004  
C:Accession: A02953  
R:Hoifmann, W.; Franz, J.K.; Franke, W.W.  
J. Mol. Biol. 184, 713-724, 1985  
A:Title: Amino acid sequence microheterogeneities of basic (type II) cytokeratins of Xenopus laevis  
A:Reference number: A92914; MUID:86011576; PMID:2413219  
A:Accession: A02953  
A:Molecule type: mRNA  
A:Residues: 1-425 <HOF>  
A:Cross-references: UNIPROT:P04266; UNIPARC:UPI000012DB06; GB:M13956; NID:G64  
A:Experimental source: clone pUP23  
C:Comment: The cytokeratin IF proteins appear to be obligate heteropolymers, composed of types of proteins is low and restricted to specific isolated regions within the rod domain. Three subdomains, each characterized by a distinctive amino acid composition, of length among the intermediate filament proteins, the arrangement of its subdomains probably distant species, suggests functional significance.  
C:Comment: There is only one group of closely related type II amphibian cytokeratins. The site of differentiation of the amphibian epidermis, which lacks a true stratum corneum.  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; intermediate filament  
F:1-296/Domain: rod (fragment) <ROD>  
F:1-16/Region: coil 1A (fragment)  
F:17-30/Region: linker 1  
F:31-131/Region: coil 1B  
F:132-148/Region: linker 12  
F:149-167/Region: coil 2A  
F:168-175/Region: linker 2  
F:176-296/Region: coil 2B  
F:234/Region: stutted  
F:297-425/Domain: tail <END>  
F:297-405/Region: H2 and V2 subdomains  
F:406-425/Region: E2 subdomain

Query Match 79.5%; Score 35; DB 1; Length 425;  
Best Local Similarity 75.0%; Pred. No. 40;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
Db 73 YEDEINKR 80

RESULT 13  
A43782  
keratin, type II - eastern newt (fragment)  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C:Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A43782  
R:Ferretti, P.; Brookes, J.P.; Brown, R.  
Development 111, 497-507, 1991  
A:Title: A new type II keratin restricted to normal and regenerating limbs and tails in the newt  
A:Reference number: A43782; MUID:91372147; PMID:1716554  
A:Accession: A43782  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-461 <FER>  
A:Cross-references: UNIPROT:Q42435; UNIPARC:UPI00001774C4; GB:X57671  
C:Superfamily: cytoskeletal keratin

Query Match 79.5%; Score 35; DB 2; Length 461;  
Best Local Similarity 75.0%; Pred. No. 43;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
Db 134 YEDEINKR 141

RESULT 14  
S05602  
keratin K7, type II, epithelial, 55K - human  
C:Species: Homo sapiens (man)

C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 29-Aug-1997  
C:Accession: S05602; A31141  
R:Glass, C.; Fuchs, E.  
J. Cell Biol. 107, 1337-1350, 1988  
A:Title: Isolation, sequence, and differential expression of a human K7 gene in simple epithelial cells  
A:Reference number: S05602; MUID:89008577; PMID:2459129  
A:Accession: S05602  
A:Molecule type: DNA  
A:Residues: 1-469 <GLA>  
A:Cross-references: UNIPARC:UPI00001774C1; EMBL:X13320  
A:Note: the authors translated the codon ACC for residue 97 as Ala  
C:Genetics:  
A:Gene: GDB:KRT7  
A:Cross-references: GDB:128114; OMIM:148059  
A:Map position: 12q12-12q21  
A:Introns: 108/3; 179/2; 179/3; 231/3; 286/3; 328/3; 402/2; 414/1  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

Query Match 79.5%; Score 35; DB 2; Length 469;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
Db 180 YEDEINRR 187

RESULT 15  
A34720  
keratin 8, type II cytoskeletal - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Jul-1990 #sequence\_revision 06-Jul-1990 #text\_change 09-Jul-2004  
C:Accession: A34720; S06888; S12479; I37982; I37983; JS0487; A31942  
R:Yamamoto, R.; Kao, L.C.; McKnight, C.E.; Strauss III, J.F.  
Mol. Endocrinol. 4, 370-374, 1990  
A:Title: Cloning and sequence of cDNA for human placental cytokeratin 8. Regulation of transcription  
A:Reference number: A34720; MUID:90259929; PMID:1692965  
A:Accession: A34720  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-483 <YAM>  
A:Cross-references: UNIPROT:P05787; UNIPARC:UPI00001774C3; GB:M34225  
R:Leube, R.E.; Bosch, F.X.; Romano, V.; Zimbelmann, R.; Hoesfler, H.; Franke, W.W.  
Differentiation 33, 69-85, 1986  
A:Title: Cytokeratin expression in simple epithelia  
A:Reference number: S06888; MUID:87134779; PMID:2434381  
A:Accession: S06888  
A:Molecule type: mRNA  
A:Residues: 205-483 <LEU>  
A:Cross-references: UNIPARC:UPI000016ABAA1; EMBL:X12882  
R:Franke, W.W.  
submitted to the EMBL Data Library, September 1988  
A:Reference number: S12479  
A:Accession: S12479  
A:Molecule type: mRNA  
A:Residues: 205-309, 'I', 311-483 <FRA>  
A:Cross-references: UNIPARC:UPI000016A769; EMBL:X12882; NID:G30312; PIDN:CAA31376.1; PID: New Biol. 2, 464-478, 1990  
R:Waseem, A.; Alexander, C.M.; Steel, J.B.; Lane, E.B.  
A:Title: Embryonic simple epithelial keratins 8 and 18: chromosomal location emphasizes a common evolutionary origin  
A:Reference number: I37982; MUID:91145351; PMID:1705144  
A:Accession: I37982  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-76, 'V', 78-416, 'S', 418-428, 'G', 430-431, 'S', 433-483 <RES>  
A:Cross-references: UNIPARC:UPI0000161B3E; EMBL:X74329; NID:G400415; PIDN:CAAS2882.1; PID: A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 151-201, 'VD', 202-271 <RE2>  
A:Cross-references: UNIPARC:UPI000016ABCF; EMBL:X74981; NID:G400417; PIDN:CAAS2916.1; PID: R:Krause, S.; Franke, W.W.

Gene 86, 241-249, 1990  
A;Title: Organization and sequence of the human gene encoding cytokeratin 8.  
A;Reference number: JS0487; MUID:90215304; PMID:1691124  
A;Accession: JS0487  
A;Molecule type: DNA  
A;Residues: 1-76, 'V', 78-428, 'G', 'SQA', 434-483 <KRA>  
A;Cross-references: UNIPARC:UPI000016A82B; GB:M34482; NID:g181572; PIDN:AAA5763.1; PID:  
R;Kulesh, D.A.; Cecena, G.; Darmon, Y.M.; Vasseur, M.; Oshima, R.G.  
Mol. Cell. Biol. 9, 1553-1565, 1989  
A;Title: Posttranslational regulation of keratins: degradation of mouse and human keratin  
A;Reference number: A31942; MUID:89261783; PMID:2471065  
A;Accession: A31942  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-76, 'V', 78-231 <KUL>  
A;Cross-references: UNIPARC:UPI000016A354; GB:M26512; NID:g177796; PIDN:AAA51542.1; PID:  
C;Comment: This protein is one of the type II cytokeratins.  
C;Genetics:  
A;Gene: GDB:KRT8  
A;Cross-references: GDB:118830; OMIM:148060  
A;Map position: 12p13.2-12q24.1  
A;Introns: 108/3; 178/2; 198/3; 230/3; 327/3; 401/2; 421/1  
A;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; intermediate filament

Query Match 79.5%; Score 35; DB 2; Length 483;  
Best Local Similarity 75.0%; Pred. NO. 45;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8

DB 179 YEDEINKR 186

Search completed: July 5, 2006, 19:20:25  
Job time : 10.7087 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 81.8641 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-11  
Perfect score: 44  
Sequence: 1 YEDEINNK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	93.2	777	Q54EF1_DICDI	Q54ef1 dictyosteli
2	40	90.9	437	Q4XNH7_PLACH	Q4xnh7 plasmodium
3	39	88.6	223	Q8X158_CLOPE	Q8x158 clostridium
4	38	86.4	171	Q4YIN2_PLABE	Q4yin2 plasmodium
5	38	86.4	541	Q4YN40_PLABE	Q4yn40 plasmodium
6	38	86.4	548	Q4LOR7_MAIZE	Q4lor7 zea mays (m
7	38	86.4	548	Q5G360_MAIZE	Q5g360 zea mays (m
8	37	84.1	121	Q33PE6_9GAMM	Q33pe6 shewanella
9	37	84.1	321	Q5UR43_MIMIV	Q5ur43 mimivirus
10	37	84.1	484	Y1062_METUA	Q58462 methanococ
11	37	84.1	506	Q96560_HELTU	Q96560 helianthus
12	37	84.1	549	Q54H37_DICDI	Q54h37 dictyosteli
13	37	84.1	559	Q8XJV8_CLOPE	Q8xjv8 clostridium
14	37	84.1	683	Q24424_PARSO	Q24424 papaver som
15	37	84.1	692	Q39035_ARATH	Q39035 arabidopsis
16	37	84.1	692	Q43235_VICSA	Q43235 vicia sativ
17	37	84.1	692	Q9AU08_9ROSI	Q9au08 populus tri
18	37	84.1	692	Q8VB48_ARATH	Q8vb48 arabidopsis
19	37	84.1	703	Q8VX49_WHEAT	Q8vx49 triticum ae
20	37	84.1	712	Q9AU06_9ROSI	Q9au06 populus tri
21	37	84.1	712	Q9AU07_9ROSI	Q9au07 populus tri
22	37	84.1	714	Q653S9_ORYSA	Q653s9 oryza sativ
23	37	84.1	1191	Q8IL10_PLAF7	Q8il10 plasmodium
24	36	81.8	372	Q9QAM2_WSSV	Q9qam2 white-spot
25	36	81.8	403	Q9YV17_MSEPV	Q9yv17 melanoplus
26	36	81.8	471	Q5HU43_CAMWR	Q5hu43 campylobact
27	36	81.8	570	Q6A1F5_DESPFS	Q6a1f5 desulfotale
28	36	81.8	595	Q663A8_BORGA	Q663a8 borrelia ga
29	36	81.8	597	Q51037_BORBU	Q51037 borrelia bu
30	36	81.8	627	Q6EBA6_CAMTE	Q6eba6 campylobact
31	36	81.8	725	Q934J5_VIBPA	Q934j5 vibrio para

32	36	81.8	725	2	Q87HI4_VIBPA	Q87hi4 vibrio para
33	36	81.8	882	2	Q4HDG9_CAMCO	Q4hdg9 campylobact
34	36	81.8	1410	2	Q7YYV2_CRYPV	Q7yyv2 cryptospori
35	36	81.8	3429	2	Q8IBP1_PLAF7	Q8ibp1 plasmodium
36	35	79.5	136	2	Q4Z5W3_PLABE	Q4z5w3 plasmodium
37	35	79.5	156	2	Q3KZ86_SCHJA	Q3kz86 schistosoma
38	35	79.5	173	2	Q35T43_GEONG	Q35t43 geobacter m
39	35	79.5	176	2	Q4J123_ICTPU	Q4j123 ictalurus p
40	35	79.5	194	2	Q4G296_TRIVA	Q4g296 trichomonas
41	35	79.5	214	2	Q7RGN1_PLAYO	Q7rgn1 plasmodium
42	35	79.5	222	2	Q4LOE7_MOUSE	Q4loe7 mus musculu
43	35	79.5	224	2	Q678L1_MOUSE	Q678l1 mus musculu
44	35	79.5	233	2	Q5DDV3_SCHJA	Q5ddv3 schistosoma
45	35	79.5	233	2	Q2WVY4_CLOBE	Q2wvy4 clostridium

## ALIGNMENTS

RESULT 1  
Q54EF1\_DICDI PRELIMINARY; PRT; 777 AA.  
ID Q54EF1\_DICDI AC Q54EF1;  
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.  
DT 24-MAY-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Hypothetical protein.  
GN ORFNames=DD80183975;  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
OX NCBI\_TaxID=44889;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=AX4;  
RX PubMed=15875012; DOI=10.1038/nature03481;  
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.A., Xu Q.,  
RA Sugang R., Berriman M., Song J., Olsen R., Konfortov B.A., Rivero F.,  
RA Tungal B., Kummerfeld S., Madera M., Konfortov B.A., Rivoir P.,  
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,  
RA Pilcher K., Chen G., Saunders D., Sodergren E.J., Davis P.,  
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,  
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.M.,  
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,  
RA Muzny D.M., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,  
RA Hauser H., James K.D., Quiles M., Madan Babu M., Saito T.,  
RA Buchrieser C., Wardrop A., Felder M., Thangavelu M., Johnson D.,  
RA Knights A., Loulèsed H., Mungall K.L., Oliver K., Price C.,  
RA Quail M.A., Urushihara H., Hernandez J., Rabinowitsch E., Steffen D.,  
RA Sanders M., Ma J., Kohara Y., Sharp S., Simmonds M.N., Spiegler S.,  
RA Tivey A., Sugano S., White B., Walker D., Woodward J.R., Winckler T.,  
RA Tanaka Y., Shaulsky G., Schleicher M., Weinstein G.M., Rosenthal A.,  
RA Cox E.C., Chisholm R.L., Gibbs R.A., Loomis W.F., Platzer M.,  
RA Kay R.R., Williams J.G., Dear P.H., Noegel A.A., Barrell B.G.,  
RA Kuspa A.;  
RT "The genome of the social amoeba Dictyostelium discoideum";  
RL Nature 435:43-57(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.

-----  
Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
Distributed under the Creative Commons Attribution-NonCommercial license  
-----  
EMBL: AAF101000257; EAL61761.1; -; Genomic\_DNA.  
DR InterPro: IPR006089; Acyl CoA DH.  
DR InterPro: IPR004813; Tetrapept\_transpt.  
DR Pfam: PF03169; OPT; 1.  
DR TIGRfams: TIGR00728; OPT\_sfam; 1.  
DR PROSITE: PS00073; ACYL\_COA\_DH\_2; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 777 AA; 85050 NW; 96BE7A19BE697E24 CRC64;  
Query Match 93.2%; Score 41; DB 2; Length 777;

```
Best Local Similarity 87.5%; Pred. No. 68;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8
Db 201 YDDEINNK 208

RESULT 2
Q4XNH7_PLACH PRELIMINARY; PRT; 437 AA.
AC Q4XNH7;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein (Fragment).
GN ORFNames=PC00683.04.0;
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15637271; DOI=10.1126/science.1103717;
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.M.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R. III,
RA Kafatos F.C., Janse C.J., Barrell B.G., Turner C.M.R., Waters A.P.,
RA Sinden R.S.;
RA "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAJ01004418; CAH81535.1; -; Genomic_DNA.
KW Hypothetical protein.
FT NON TER 1
FT NON TER 437
FT NON TER 437
SQ SEQUENCE 437 AA; 51115 MW; 45720BA01496B79A CRC64;

Query Match 90.9%; Score 40; DB 2; Length 437;
Best Local Similarity 87.5%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8
Db 179 YDEINNK 186

RESULT 3
Q8X158_CLOPE PRELIMINARY; PRT; 223 AA.
AC Q8X158;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 18.
DE Hypothetical protein CPE2263.
GN OrderedLocNames=CPE2263;
OS Clostridium perfringens
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
```

```
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
"Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BA000016; BAB81969.1; -; Genomic_DNA.
DR Biocyc; CPER1502:CPE2283-MONOMER; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00745; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Complete proteome; Hypothetical protein;
KW Peptidoglycan-anchor.
SQ SEQUENCE 223 AA; 25063 MW; F259D8A358DD968E CRC64;

Query Match 88.6%; Score 39; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINNK 7
Db 177 YEDEINNK 183

RESULT 4
Q4YIN2_PLABE PRELIMINARY; PRT; 171 AA.
AC Q4YIN2;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Hypothetical protein (Fragment).
GN ORFNames=PB300572.00.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15637271; DOI=10.1126/science.1103717;
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.M.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R. III,
RA Kafatos F.C., Janse C.J., Barrell B.G., Turner C.M.R., Waters A.P.,
RA Sinden R.S.;
RA "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAJ01004738; CAI02133.1; -; Genomic_DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007114; MFS.
DR PROSITE; PS50850; MFS; 1.
KW Hypothetical protein.
FT NON TER 171
FT NON TER 171
SQ SEQUENCE 171 AA; 18905 MW; 4F77BBA698BES146 CRC64;

Query Match 86.4%; Score 38; DB 2; Length 171;
Best Local Similarity 85.7%; Pred. No. 54;
```



```
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 YEDEINN 7
Db 33 YEDEVNN 39
|||||
RESULT 5
Q4YN40_PLABE PRELIMINARY; PRT; 541 AA.
AC Q4YN40;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 21-FEB-2006, entry version 8.
DE Hypothetical protein.
GN ORFNAMES=PB001031.03.0;
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15637271; DOI=10.1126/science.1103717;
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Bertman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.M.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R. III,
RA Kafatos F.C., Janse C.J., Barrell B.G., Turner C.M.R., Waters A.P.,
RA Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (By
CC similarity).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; CAAI01003411; CAI00570.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007114; MFS.
DR Pfam; PF07690; MFS 1; 1.
DR PROSITE; PS50850; MFS, 1.
KW Hypothetical protein; Membrane; Transmembrane; Transport.
SQ SEQUENCE 541 AA; 59919 MW; 865C51731BCFCD8 CRC64;
Query Match 86.4%; Score 38; DB 2; Length 541;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 YEDEINN 7
Db 33 YEDEVNN 39
|||||
RESULT 6
Q4LOR7_MAIZE PRELIMINARY; PRT; 548 AA.
AC Q4LOR7;
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 02-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Putative sesquiterpene cyclase.
GN Name=Umi2;
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Zea mays (Maize).
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15980197; DOI=10.1104/pp.105.061200;
RA Basse C.W.;
RT "Dissecting Defense-Related and Developmental Transcriptional
RT Responses of Maize during Ustilago maydis Infection and Subsequent
RT Tumor Formation.";
RL Plant Physiol. 138:1774-1784(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AY647253; AAT70085.1; -; mRNA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001906; Terp_synth-like.
DR InterPro; IPR005630; Tps_metal_bd.
DR Pfam; PF01397; Terpene_synth; 1.
DR Pfam; PF03936; Terpene_synth C; 1.
SQ SEQUENCE 548 AA; 63731 MW; 5A58DE8B437F7C4C CRC64;
Query Match 86.4%; Score 38; DB 2; Length 548;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 YEDEINN 8
Db 88 YEDEINEK 95
|||||
RESULT 7
Q5GJ60_MAIZE PRELIMINARY; PRT; 548 AA.
AC Q5GJ60;
DT 01-MAR-2005, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Terpene synthase 6.
GN Name=TPS6;
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15980197; DOI=10.1104/pp.105.061200;
RA Basse C.W.;
RT "Dissecting Defense-Related and Developmental Transcriptional
RT Responses of Maize during Ustilago maydis Infection and Subsequent
RT Tumor Formation.";
RL Plant Physiol. 138:1774-1784(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AY647253; AAT70085.1; -; mRNA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001906; Terp_synth-like.
DR InterPro; IPR005630; Tps_metal_bd.
DR Pfam; PF01397; Terpene_synth; 1.
DR Pfam; PF03936; Terpene_synth C; 1.
SQ SEQUENCE 548 AA; 63731 MW; 5A58DE8B437F7C4C CRC64;
Query Match 86.4%; Score 38; DB 2; Length 548;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 YEDEINN 8
Db 88 YEDEINEK 95
|||||
RESULT 8
Q33PE6_9GAMM PRELIMINARY; PRT; 121 AA.
ID Q33PE6_9GAMM
```

```
AC Q33PE6;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DE Dihydroneopterin aldolase.
GN ORFNames=ShewDRAFT_3672;
OS Shewanella sp. PV-4.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OX Shewanellaceae; Shewanella.
RX NCBI_TaxID=323850;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=PV-4;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Shewanella sp. PV-4.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=PV-4;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella sp. PV-4.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC
CC EMBL; AAL501000063; EAP00981.1; -; Genomic DNA.
DR GO; GO:0004150; F: dihydroneopterin aldolase activity; IEA.
DR GO; GO:0006760; P: folic acid and derivative metabolism; IEA.
SQ SEQUENCE 121 AA; 13946 MW; 6F808F043D66EBA1 CRC64;

Query Match 84.1%; Score 37; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EDEINNK 8
Db 23 EDEINNK 29

RESULT 9
Q5UR43_MIMIV
ID Q5UR43 MIMIV PRELIMINARY; PRT; 321 AA.
AC Q5UR43;
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 1.
DE Hypothetical protein.
GN ORFNames=MIMI_R562;
OS Mimivirus.
OC Viruses; dsDNA viruses, no RNA stage; Mimivirus.
OX NCBI_TaxID=212035;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=Rowbotham-Bradford;
RX MEDLINE=22550848; PubMed=12663918; DOI=10.1126/science.1081867;
RA La Scola B., Audic S., Robert C., Jungang L., de Lamballerie X.,
RA Drancourt M., Bittles R., Claverie J.M., Raoult D.;
RT "A giant virus in amoebae.";
RL Science 299:2033-2033 (2003).
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=Rowbotham-Bradford;
RX PubMed=15486256; DOI=10.1126/science.1101485;
RA Raoult D., Audic S., Robert C., Abergel C., Renesto P., Ogata H.,
RA La Scola B., Susan M., Claverie J.-M.;
RT "The 1.2-megabase genome sequence of Mimivirus.";

Q33PE6;
Science 306:1344-1350 (2004).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC
CC EMBL; AY653733; AAV50826.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 321 AA; 37561 MW; F5D2CE4E266EBA4B CRC64;

Query Match 84.1%; Score 37; DB 2; Length 321;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8
Db 50 YEDEFNNK 57

RESULT 10
Y1062_METJA
ID Y1062 METJA STANDARD; PRT; 484 AA.
AC Q58462;
DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1996, sequence version 1.
DT 07-MAR-2006, entry version 29.
DE Hypothetical protein MJ1062.
GN OrderedLocustNames=MJ1062;
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OX Methanocaldococcaceae; Methanocaldococcus.
RX NCBI_TaxID=2190;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=JAU-1 / DSM 2661 / ATCC 43067;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadow P.W., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073 (1996).

Q33PE6;
Science 306:1344-1350 (2004).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC
CC EMBL; L77117; AAB99065.1; -; Genomic DNA.
DR PIR; E64432; E64432.
DR GenomeReviews; L77117_GR; MJ1062.
DR TIGR; MJ1062; -.
DR BiOCyc; MJAN243232; MJ1062-MONOMER; -.
DR InterPro; IPR000182; GCN5acetyl trans.
DR Pfam; PF00583; Acetyltransf 1; I.
KW Complete proteome; Hypothetical protein.
FT CHAIN 1 484 Hypothetical protein MJ1062.
FT /FTID=PRO_0000107155.
SQ SEQUENCE 484 AA; 56694 MW; E1764675A97CFEB2 CRC64;

Query Match 84.1%; Score 37; DB 1; Length 484;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7
Db 211 YEDELNN 217

RESULT 11
Q96560_HELTU
```

```

ID Q96560_HELTU PRELIMINARY; PRT; 506 AA.
AC Q96560;
DT 01-FEB-1997, integrated into UniProtKB/TrEMBL.
DT 01-FEB-1997, sequence version 1.
DE NADPH-ferrihemoprotein reductase (EC 1.6.2.4) (Fragment).
OS Helianthus tuberosus (Jerusalem artichoke).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC asterids; campanulids; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4233;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=blanc commun; TISSUE=Parenchyma;
RA Hasenfratz M.;
RL Thesis (1992); Cellular and Molecular Enzymology, CNRS - Institut de
RL Biologie Moleculaire des Plantes.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=blanc commun; TISSUE=Parenchyma;
RA Hasenfratz M., Jeitsch J., Benveniste I., Lesot A., Durst F.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=blanc commun; TISSUE=Parenchyma;
RA Benveniste I.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC
DR EMBL; Z26251; CAA81210.1; -; mRNA.
DR PIR; S37156; S37156.
DR HSP; P00388; 1A00.
DR GO; GO:0010181; F-FMN binding; IEA.
DR GO; GO:0003958; F-NADPH-hemoprotein reductase activity; IEA.
DR GO; GO:0016491; F-oxidoreductase activity; IEA.
DR GO; GO:0006118; P-electron transport; IEA.
DR InterPro; IPR003097; FAD bd.
DR InterPro; IPR008254; Flavodoxin 1.
DR InterPro; IPR001709; FPN cyt reductase.
DR InterPro; IPR001433; Oxid FAD NAD bd.
DR Pfam; PF00667; FAD_binding_1; 1.
DR Pfam; PF00258; Flavodoxin_1; 1.
DR Pfam; PF00175; NAD binding_1; 1.
DR PRINTS; PR00371; FPNCR.
DR PROSITE; PS0902; FLAVODOXIN_LIKE; 1.
KW Oxidoreductase.
FT NON_TER 1
SQ SEQUENCE 506 AA; 56229 MW; 05F34FF248E82B08 CRC64;

Query Match 84.1%; Score 37; DB 2; Length 506;
Best Local Similarity 85.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7
DB 405 YEDELNN 411

RESULT 12
ID Q54H37_DICDI PRELIMINARY; PRT; 549 AA.
AC Q54H37;
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 24-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN ORFNames=DPB0188544;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;

```

```

RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RX PubMed=15875012; DOI=10.1038/nature03481;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.A., Xu Q.,
RA Sugchang R., Berriman M., Song J., Olsen R., Szafranski K., Rivero P.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivoire P.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E.J., Davis P.,
RA Karhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Perbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.M.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D.M., Mourier T., Pain A., Lu M., Harper D., Lindsey R.,
RA Hauser H., James K.D., Quiles M., Madan Babu M., Saito T.,
RA Buchrieser C., Wardrop A., Felder M., Thangavelu M., Johnson D.,
RA Knights A., Loulseghe H., Mungall K.L., Oliver K., Price C.,
RA Quail M.A., Urushihara H., Hernandez J., Rabinowitsch E., Steffen D.,
RA Sanders M., Ma J., Kohara Y., Sharp S., Simmonds M.N., Spiegler S.,
RA Tivey A., Sugano S., White B., Walker D., Woodward J.R., Winckler T.,
RA Tanaka Y., Shaalsky G., Schleicher M., Weinstock G.M., Rosenthal A.,
RA Cox E.C., Chisholm R.L., Gibbs R.A., Loomis W.F., Platzer M.,
RA Kay R.R., Williams J.G., Dear P.H., Noegel A.A., Barrell B.G.,
RA Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 435:43-57(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC
DR EMBL; AAFI01000219; EALG2574.1; -; Genomic_DNA.
DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0004930; F-G-protein coupled receptor activity; IEA.
DR InterPro; IPR000832; GPCR secretin.
DR PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW Hypothetical protein.
SQ SEQUENCE 549 AA; 63275 MW; 30F3E2E37B4E1280 CRC64;

Query Match 84.1%; Score 37; DB 2; Length 549;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EDEINN 8
DB 531 EDEINN 537

RESULT 13
Q8XJVB_CLOPE PRELIMINARY; PRT; 559 AA.
ID Q8XJVB_CLOPE PRELIMINARY; PRT; 559 AA.
AC Q8XJVB;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 18.
DE Mg-protocorphyrin IX monomethyl ester oxidative cyclase.
GN OrderedLocustNames=CPE1645;
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms

```

```
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BA000016; BAB81351.1; -; Genomic DNA.
DR BiOcyC; CPER1502:CPE1645-MONOMER; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0050897; F:cobalt ion binding; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR InterPro; IPR006158; B12 bd.
DR InterPro; IPR006838; Eip3/MiaB/NifB.
DR InterPro; IPR007197; Radical SAM.
DR InterPro; IPR005839; UPF0004.
DR PANTHER; PTHR11918; UPF0004; 1.
DR Pfam; PF02310; B12-binding; 1.
DR Pfam; PF04055; Radical SAM; 1.
DR SMART; SM00729; Elp3; 1.
KW Complete proteome.
SQ SEQUENCE 559 AA; 66243 MW; ED3739DB822BBDF8 CRC64;

Query Match 84.1%; Score 37; DB 2; Length 559;
Best Local Similarity 75.0%; Pred No. 2.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 8
Db 166 YEEDINN 173

RESULT 14
O24424_PAPSO
ID O24424_PAPSO PRELIMINARY; PRT; 683 AA.
AC O24424;
DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.
DT 01-JAN-1998, sequence version 1.
DT 07-FEB-2006, entry version 29.
DE NADPH:ferrihemoprotein oxidoreductase (EC 1.6.2.4).
OS Papaver somniferum (Opium poppy).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
OC Papaveraceae; papaver.
OX NCBI_TaxID=3469;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98096363; PubMed=9434750; DOI=10.1006/abbi.1997.0374;
RA Rosco A., Pauli H.H., Priesner W., Kutchan T.M.;
RT "Cloning and heterologous expression of NADPH-cytochrome P450
RT reductases from the Papaveraceae.";
RL Arch. Biochem. Biophys. 348:369-377 (1997).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; U67185; AAC05021.1; -; mRNA.
DR FIR; T10720; T10720.
DR HSP; P00388; IAMO.
DR GO; GO:0010181; F:FMN binding; IEA.
DR GO; GO:0003958; F:NADPH-hemoprotein reductase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR003097; FAD bd.
DR InterPro; IPR001094; Flavodoxin like.
DR InterPro; IPR008254; Flavodoxin 1.
DR InterPro; IPR001709; FPN cyt reductase.
DR InterPro; IPR001433; Oxred_FAD_NAD_bd.
DR Pfam; PF00657; FAD binding_1; 1.
DR Pfam; PF00258; Flavodoxin_1; 1.
DR PRINTS; PR00175; NAD binding_1; 1.
DR PRINTS; PR00371; FPNCR.
DR PROSITE; PS0902; FLAVODOXIN_LIKE; 1.
KW Oxidoreductase.
SQ SEQUENCE 683 AA; 75888 MW; E07BD5E5EA8DE002 CRC64;

Query Match 84.1%; Score 37; DB 2; Length 683;
Best Local Similarity 85.7%; Pred No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7
Db 591 YEDELNN 597

Search completed: July 5, 2006, 19:01:22
Job time : 83.9641 secs

Best Local Similarity 85.7%; Pred No. 3.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7
Db 582 YEDELNN 588

RESULT 15
Q39035_ARATH
ID Q39035_ARATH PRELIMINARY; PRT; 692 AA.
AC Q39035;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 31.
DE NADPH:ferrihemoprotein reductase.
GN Name=ATRI;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsia.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Heynh; TISSUE=Whole seedling;
RA Mignote-Vieux C., Kazmaier M., Lacroute F., Pompon D.M.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; X66016; CAA46814.1; -; mRNA.
DR HSP; P00388; IAMO.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0010181; F:FMN binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR003097; FAD bd.
DR InterPro; IPR001094; Flavodoxin like.
DR InterPro; IPR008254; Flavodoxin 1.
DR InterPro; IPR001709; FPN cyt reductase.
DR InterPro; IPR001433; Oxred_FAD_NAD_bd.
DR Pfam; PF00657; FAD binding_1; 1.
DR Pfam; PF00258; Flavodoxin_1; 1.
DR PRINTS; PR00175; NAD binding_1; 1.
DR PRINTS; PR00371; FPNCR.
DR PROSITE; PS0902; FLAVODOXIN_LIKE; 1.
SQ SEQUENCE 692 AA; 76769 MW; 5E0216F008B5D49E CRC64;

Query Match 84.1%; Score 37; DB 2; Length 692;
Best Local Similarity 85.7%; Pred No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7
Db 591 YEDELNN 597

Search completed: July 5, 2006, 19:01:22
Job time : 83.9641 secs
```

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 69.5146 Seconds  
(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928A-11  
Perfect score: 44  
Sequence: 1 YEDEINNK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 8:\*

1:	Geneseqp1980s:*
2:	Geneseqp1990s:*
3:	Geneseqp2000s:*
4:	Geneseqp2001s:*
5:	Geneseqp2002s:*
6:	Geneseqp2003as:*
7:	Geneseqp2003bs:*
8:	Geneseqp2004s:*
9:	Geneseqp2005s:*
10:	Geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	44	100.0	8	6	AAO26391	Psoriasis	
2	44	100.0	8	9	ADVO4416	AadV04416 Leishmani	
3	37	84.1	683	2	AAW85682	Poppy cyt	
4	37	84.1	692	2	AAR43581	Plant NAD	
5	37	84.1	698	9	AEA34225	Opium pop	
6	37	84.1	703	8	ADT59926	Plant pol	
7	37	84.1	883	2	AAW85680	Poppy cyt	
8	36	81.8	120	5	ABP42597	Human ova	
9	36	81.8	217	4	AAE87629	Bovine ma	
10	36	81.8	368	4	AAG85031	Shrimp wh	
11	36	81.8	370	8	ADX94912	Plant ful	
12	36	81.8	407	8	ADT57158	Plant pol	
13	36	81.8	469	8	ABM81385	Tumour-as	
14	36	81.8	546	3	AAV52397	Human ker	
15	36	81.8	546	5	AAE20422	Human ker	
16	35	79.5	113	6	ABG74871	Human cyt	
17	35	79.5	132	8	ADR98929	Lung spec	
18	35	79.5	139	2	AAW18011	Plasmodiu	
19	35	79.5	180	8	ADN99441	Novel hum	
20	35	79.5	186	4	AAW16318	Peptide #	
21	35	79.5	186	4	ABB35310	Peptide #	
22	35	79.5	186	4	AAW28813	Peptide #	
23	35	79.5	186	4	ABB30140	Peptide #	

24	35	79.5	186	4	ABB20752	Protein #
25	35	79.5	186	4	AAW68514	Human bon
26	35	79.5	186	4	AAW56141	Human bra
27	35	79.5	186	4	ABG50183	Human liv
28	35	79.5	186	4	AAW04056	Peptide #
29	35	79.5	186	4	ABG09410	Novel hum
30	35	79.5	186	5	ABG38095	Human pep
31	35	79.5	233	8	ADR98930	Lung spec
32	35	79.5	236	5	ABP43147	Human ova
33	35	79.5	250	8	ADQ67770	Novel hum
34	35	79.5	252	8	ADR10191	Human pro
35	35	79.5	273	4	ABG26376	Novel hum
36	35	79.5	275	4	ABG15282	Novel hum
37	35	79.5	282	8	ABM80614	Tumour-as
38	35	79.5	342	7	ADE15641	Human str
39	35	79.5	398	6	AAE38086	Human cyt
40	35	79.5	401	8	ABM80365	Tumour-as
41	35	79.5	403	8	ABM81640	Tumour-as
42	35	79.5	420	8	ADP29879	Human sec
43	35	79.5	429	4	ABG19547	Novel hum
44	35	79.5	429	4	ABG19546	Novel hum
45	35	79.5	433	3	AAW58755	Breast an

ALIGNMENTS

RESULT 1	AAO26391	AAO26391 standard; peptide; 8 AA.
XX	XX	
AC	AAO26391;	
XX	XX	
DT	30-JAN-2003 (first entry)	
XX	XX	Psoriasis treating immunotherapeutic peptide, SEQ ID No 11.
DE	XX	
XX	XX	Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;
KW	XX	vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;
KW	XX	gene therapy.
XX	XX	
OS	XX	Leishmania sp.
XX	XX	
PN	XX	WO200274239-A2.
XX	XX	
PD	XX	26-SEP-2002.
XX	XX	
PF	XX	04-MAR-2002; 2002WO-US006496.
XX	XX	
PR	XX	16-MAR-2001; 2001US-00809003.
XX	XX	(AKIV-) AKIVA LLC.
PA	XX	O'daly JA;
PI	XX	
XX	XX	WPI; 2003-018763/01.
DR	XX	
XX	XX	New polypeptide and immunogenic variants comprising amino acid sequences
PT	XX	of particulate antigens, useful for the treatment and clinical remission
PT	XX	of psoriasis.
PT	XX	
XX	XX	Claim 1; Page 43; 56pp; English.
PS	XX	
XX	XX	The invention relates to a polypeptide comprising an isolated amino acid
CC	XX	sequence or immunogenic variants selected from any of 14 fully defined
CC	XX	sequences of 7-16 amino acids, given in the specification. The
CC	XX	immunotherapeutic agents and a pharmaceutical compositions comprising
CC	XX	polynucleotides and vectors of the invention are useful for the treatment
CC	XX	and clinical remission of psoriasis. The isolated nucleic acids are
CC	XX	useful as probes. The sequences of the invention can be used in the
CC	XX	treatment of disorders by gene therapy. This sequence represents one of
CC	XX	the 14 immunotherapeutic peptides of the invention

SQ Sequence 8 AA;

Query Match 100.0%; Score 44; DB 6; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8  
| | | | | | | |  
Db 1 YEDEINNK 8

RESULT 2  
ADV04416  
ID ADV04416 standard; peptide; 8 AA.  
XX  
AC ADV04416;  
XX  
DT 24-FEB-2005 (first entry)  
XX  
DE Leishmania immunotherapeutic peptide SEQ ID NO:11.  
XX  
KW T-lymphocyte; immunogenicity; antigen; antiprosoriatric; immunostimulant;  
KW psoriasis.  
XX  
OS Leishmania sp.  
XX  
PN US2004241168-A1.  
XX  
PD 02-DEC-2004.  
XX  
PF 09-FEB-2004; 2004US-00774928.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
PR 17-OCT-2003; 2003US-00687892.  
XX  
PA (ODAL/) ODALY J A.  
XX  
PI Odaly JA;  
XX  
DR WPI; 2005-011563/01.  
XX  
PT Inhibiting selectively T-cell rolling in human, by administering compound  
PT interfering with cutaneous lymphocyte antigen-E selectin interaction and  
PT leukocyte function associated antigen-1 interaction.  
XX  
PS Disclosure; SEQ ID NO 11; 21pp; English.  
XX  
CC The invention relates to a novel method for selectively inhibiting T-cell  
CC rolling in a human host, comprising administering a compound that  
CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E  
CC selectin interaction and leukocyte function associated antigen (LFA) -  
CC 1/intercellular adhesion molecule (ICAM) and very late antigen  
CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound  
CC includes an immunotherapeutic agent, which comprises a purified protein  
CC extract that is isolated by diethylaminoethyl Sephadex chromatography of  
CC a Nonidet P-40 insoluble particulate antigen fraction derived from  
CC isolated killed cells of anastigotes from one or more species of the  
CC Leishmania genus, where the particulate antigen fraction is solubilized  
CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied  
CC to diethylaminoethyl Sephadex and is eluted with a solution comprising  
CC 0.1 M sodium chloride, 8 M urea and 0.025 M  
CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract  
CC includes polypeptides having apparent molecular weights of 73, 80 and 82  
CC kDa, after total reduction and alkylation. The species is Leishmaniosis,  
CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the  
CC invention has antiprosoriatric and immunostimulant activity, and inhibits T-  
CC cell rolling by interfering with CLA-E selectin interaction and LFA-  
CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively  
CC inhibiting T-cell rolling in a human host, and for treating psoriasis.  
CC The present sequence represents a Leishmania peptide used in the  
CC invention.  
XX  
SQ Sequence 8 AA;

Query Match 100.0%; Score 44; DB 9; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8  
| | | | | | | |  
Db 1 YEDEINNK 8

RESULT 3  
AAW85682  
ID AAW85682 standard; protein; 683 AA.  
XX  
AC AAW85682;  
XX  
DT 19-JUL-1999 (first entry)  
XX  
DE Poppy cytochrome P450 reductase.  
XX  
KW Opium poppy; alkaloid; cytochrome P450 reductase; morphine; codeine;  
KW oripavine; thebaine; transformation; crop yield; probe; primer.  
XX  
OS Papaver somniferum.  
XX  
PN WO9911765-A1.  
XX  
PD 11-MAR-1999.  
XX  
PF 28-AUG-1998; 98WO-AU000705.  
XX  
PR 29-AUG-1997; 97AU-00008872.  
XX  
PA (JOHJ ) JOHNSON & JOHNSON RES PTY LTD.  
XX  
PI Kutchan TW, Zenk MH, Atkins DG, First AJ;  
XX  
DR WPI; 1999-214703/18.  
DR N-PSDB; AAX08517.  
XX  
PT Nucleic acid encoding cytochrome P-450 reductase from poppy.  
XX  
PS Claim 19; Fig 10a; 58pp; English.  
XX  
CC Transforming plants with a nucleic acid molecule encoding cytochrome P450  
CC reductase alters the yield and/or type of alkaloids produced. In opium  
CC poppies it specifically increases the yield of medically useful alkaloids  
CC such as morphine, codeine, oripavine and thebaine. Sequences  
CC complementary to the coding sequence of cytochrome P450 reductase are  
CC useful as probes, primers and antisense sequences, or for design of  
CC ribozymes. Transformation with DNA encoding the cytochrome P450 reductase  
CC allows regulation of the total alkaloid content and of the relative  
CC proportions of individual alkaloids. Produced. Increasing the alkaloid  
CC content of poppy straw should reduce the cost of alkaloid production.  
CC Cytochrome P450 reductase is the rate-limiting enzyme in biosynthesis of  
CC alkaloids in poppies. Peptide fragments of the cytochrome P450 reductase  
CC are described in AAW85672-W85678  
XX  
SQ Sequence 683 AA;

Query Match 84.1%; Score 37; DB 2; Length 683;  
Best Local Similarity 85.7%; Pred. No. 3.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7  
| | | | | | | |  
Db 582 YEDELNN 588

RESULT 4  
AAR43581  
ID AAR43581 standard; protein; 692 AA.  
XX

AC AAR43581;  
XX  
DT 25-MAR-2003 (revised)  
DT 11-MAY-1994 (first entry)  
XX  
DE Plant NADPH cytochrome P450 reductase (ara B).  
XX  
XX NADPH cytochrome P450 reductase; functional complementation;  
KW identification.  
XX  
XX Arabidopsis thaliana.  
XX  
PN WO9321326-A2.  
XX  
XX 28-OCT-1993.  
XX  
XX 13-APR-1993; 93WO-FR000367.  
PF  
XX 13-APR-1992; 92FR-00004491.  
PR  
XX (ORSA-) ORSAN.  
XX  
XX Kazmaier M, Lacroute F, Mignotte-Vieux C, Minet M, Pompon D;  
PI WPI; 1993-351736/44.  
XX  
DR N-PSDB; AAQ51236.  
DR  
XX New DNA encoding plant NADPH cytochrome P450 reductase - cloned by  
PT functional complementation in yeast, also recombinant enzyme useful in  
PT P450 mediated bioconversion processes.  
XX  
XX Claim 13; Fig 9; 79pp; French.  
PS  
XX A new method for determining whether a DNA sequence encodes an NADPH  
CC cytochrome P450 reductase involves transforming yeasts with plasmids of a  
CC total cDNA bank of plant(s). The yeasts used in the procedure are  
CC incapable of producing their own NADPH cytochrome P450 reductase. They  
CC are then exposed to a cytochrome P450 inhibitor at a level which is  
CC lethal to the yeast cells but not to cells which, because of the  
CC transformation, now contain an active NADPH cytochrome P450 reductase.  
CC Surviving clones are then isolated and plasmid DNA extracted. The gene is  
CC inserted into the plasmid at a site which places it under the control of  
CC an inducible promoter. (Updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 692 AA;  
Query Match 84.1%; Score 37; DB 2; Length 692;  
Best Local Similarity 85.7%; Pred. No. 3.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 YEDEINN 7  
Db 591 YEDELNN 597  
RESULT 5  
AEA34225  
ID AEA34225 standard; protein; 698 AA.  
XX  
AC AEA34225;  
XX  
XX 25-AUG-2005 (first entry)  
DT  
XX Opium poppy cytochrome P450 reductase protein.  
DE  
XX Cytochrome P450 reductase; enzyme; enzyme engineering.  
KW  
XX Papaver somniferum.  
OS  
XX WO2005053382-A1.  
PN  
XX 16-JUN-2005.  
PD  
XX

PF 02-DEC-2004; 2004WO-AU001699.  
XX  
PR 02-DEC-2003; 2003AU-00906672.  
XX  
PA (JOHU ) JOHNSON & JOHNSON RES PTY LTD.  
XX  
PI Allen R, Larkin P;  
XX  
DR WPI; 2005-435265/44.  
DR N-PSDB; AEA34224.  
XX  
PT Novel polynucleotide encoding cytochrome P-450 enzyme of alkaloid  
PT producing poppy plant and inhibiting accumulation of benzyloquinoline  
PT alkaloid, useful for producing poppy plant with altered alkaloid content.  
XX  
PS Example 1; SEQ ID NO 2; 67pp; English.  
XX  
XX The invention relates to a polynucleotide encoding a cytochrome P-450  
CC reductase enzyme of an alkaloid producing poppy plant involved in a step  
CC in an alkaloid biosynthesis pathway of the plant and which inhibits  
CC accumulation of a benzyloquinoline alkaloid above a normal level of the  
CC alkaloid in the plant, or which encodes an active fragment, homolog or  
CC its variant having enzymatic activity of the enzyme. The invention also  
CC relates to a recombinant vector incorporating the polynucleotide, a  
CC purified cytochrome P-450 reductase enzyme encoded by the polynucleotide  
CC or its active fragment, homolog, analog, derivative or variant having  
CC enzymatic activity of the enzyme, and a method of providing an alkaloid  
CC producing poppy plant with an altered alkaloid content involving up-  
CC regulating or down-regulating the activity of the reductase enzyme. The  
CC polynucleotide is useful for producing an alkaloid producing poppy plant  
CC with altered alkaloid content and for providing a transgenic poppy plant  
CC altered alkaloid content. This sequence represents the opium poppy  
CC cytochrome P-450 reductase enzyme polypeptide of the invention.  
XX  
SQ Sequence 698 AA;  
Query Match 84.1%; Score 37; DB 9; Length 698;  
Best Local Similarity 85.7%; Pred. No. 3.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 YEDEINN 7  
Db 597 YEDELNN 603  
RESULT 6  
ADT59926  
ID ADT59926 standard; protein; 703 AA.  
XX  
AC ADT59926;  
XX  
DT 13-JAN-2005 (first entry)  
DT  
XX Plant polypeptide, SEQ ID 10003.  
DE  
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX  
OS Viridiplantae.  
XX  
XX US2004216190-A1.  
PN  
XX 28-OCT-2004.  
PD  
XX 18-DEC-2003; 2003US-00739930.  
PF  
XX 28-APR-2003; 2003US-00424599.  
PR  
XX 28-APR-2003; 2003US-00425115.  
XX  
XX (KOVA/) KOVALIC D K.  
PA

XX FI Kovalic DK;  
XX DR WPI; 2004-757369/74.  
XX PT New recombinant DNA constructs useful in the field of biochemistry and  
XX PT genetics, and in particular for producing transgenic plants with improved  
XX PT biological characteristics.  
XX PS Claim 2; SEQ ID NO 10003; 14pp; English.  
XX CC The invention relates a recombinant DNA construct comprising a  
XX CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
XX CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
XX CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
XX CC Arabidopsis, wheat and rape but the specification does not indicate which  
XX CC sequences is derived from which organism. Also included is a method of  
XX CC producing a plant having an improved property, comprising transforming a  
XX CC plant with a recombinant DNA construct comprising a promoter region  
XX CC functional in a plant cell operably joined to a polynucleotide encoding a  
XX CC polypeptide associated with the property, and growing the transformed  
XX CC plant. The property is selected from improving plant cold tolerance, for  
XX CC manipulating growth rate in plant cells by modification of the cell cycle  
XX CC pathway, for improving plant drought tolerance, for providing increased  
XX CC resistance to plant disease, for galactomanan production, for production  
XX CC of plant growth regulators, for improving plant heat tolerance, for  
XX CC improving plant tolerance to herbicides, for increasing the rate of  
XX CC homologous recombination in plants, for lignin production, for improving  
XX CC plant tolerance to extreme osmotic conditions, for improving plant  
XX CC tolerance to pathogens or pests, for yield improvement by modification of  
XX CC photosynthesis, for modifying seed oil yield and/or content, for  
XX CC modifying seed protein yield and/or content, for yield improvement by  
XX CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
XX CC and for yield improvement by providing improved plant growth and  
XX CC development under at least one stress condition. The polynucleotide may  
XX CC also encode a plant transcription factor. The methods and compositions of  
XX CC the present invention are useful in the field of biochemistry and  
XX CC genetics, in particular for producing transgenic plants with improved  
XX CC biological characteristics such as increased yield, improved nitrogen  
XX CC flow, increasing plant tolerance to cold or heat, improving plant  
XX CC tolerance to extreme osmotic and drought conditions, and improving plant  
XX CC tolerance to plant pests or pathogens. They can also be used in physical  
XX CC arrays of molecules, plant breeding markers, computer-based storage and  
XX CC analysis systems. The present sequence is one of the 5544 plant protein  
XX CC sequences of the invention. Note: The sequence data for this patent did  
XX CC not form part of the printed specification, but was obtained in  
XX CC electronic format directly from USPTO at  
XX CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX SQ Sequence 703 AA;

Query Match 84.1%; Score 37; DB 8; Length 703;  
Best Local Similarity 85.7%; Pred. No. 3.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
||||:|  
DB 250 YEDELNN 256

RESULT 7  
AAW85680  
ID AAW85680 standard; protein; 883 AA.

XX AC AAW85680;

XX DT 19-JUL-1999 (first entry)

XX DE Poppy cytochrome P450 reductase.

XX KW Opium poppy; alkaloid; cytochrome P450 reductase; morphine; codeine;  
XX KW oripavine; thebaine; transformation; crop yield; probe; primer.

OS Papaver somniferum.  
XX FH Key Location/Qualifiers  
XX FT Protein 42..746  
XX FT /label= Cytochrome P450 reductase  
XX PN WO9911765-A1.  
XX XX 11-MAR-1999.  
XX PD 28-AUG-1998; 98WO-AU000705.  
XX PF 29-AUG-1997; 97AU-00008872.  
XX PR (JOHJ ) JOHNSON & JOHNSON RES PTY LTD.  
XX PA Kutchan TM, Zenk MH, Atkins DG, Fist AJ;  
XX PI WPI; 1999-214703/18.  
XX DR N-PSDB; AAX08517.  
XX DR Nucleic acid encoding cytochrome P-450 reductase from poppy.  
XX PT Claim 19; Fig 9a; 58pp; English.  
XX CC Transforming plants with a nucleic acid molecule encoding cytochrome P450  
XX CC reductase alters the yield and/or type of alkaloids produced in opium  
XX CC poppies it specifically increases the yield of medically useful alkaloids  
XX CC such as morphine, codeine, oripavine and thebaine. Sequences  
XX CC complementary to the coding sequence of cytochrome P450 reductase are  
XX CC useful as probes, primers and antisense sequences, or for design of  
XX CC ribozymes. Transformation with DNA encoding the cytochrome P450 reductase  
XX CC allows regulation of the total alkaloid content and of the relative  
XX CC proportions of individual alkaloids. Produced. Increasing the alkaloid  
XX CC content of poppy straw should reduce the cost of alkaloid production.  
XX CC Cytochrome P450 reductase is the rate-limiting enzyme in biosynthesis of  
XX CC alkaloids in poppies. Peptide fragments of the cytochrome P450 reductase  
XX CC are described in AAW85672-W85678  
XX SQ Sequence 883 AA;

Query Match 84.1%; Score 37; DB 2; Length 883;  
Best Local Similarity 85.7%; Pred. No. 4.8e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
||||:|  
DB 623 YEDELNN 629

RESULT 8  
ABP42597  
ID ABP42597 standard; protein; 120 AA.

XX AC ABP42597;

XX DT 22-AUG-2002 (first entry)

XX DE Human ovarian antigen HOGCG83, SEQ ID NO:3729.

XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
XX KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
XX KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
XX KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
XX KW inflammatory condition; immune disorder; blood disorder;  
XX KW cardiovascular disorder; respiratory disorder; neurological disorder;  
XX KW gastrointestinal disorder; urinary system disorder; drug screening;  
XX KW gene therapy; chromosome mapping; forensic analysis;  
XX KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
XX KW antiinflammatory; gynaecological; reproductive.

XX OS Homo sapiens.



PN WO200200677-A1.  
XX  
PD 03-JAN-2002.  
XX  
PF 07-JUN-2001; 2001WO-US018569.  
XX  
PR 07-JUN-2000; 2000US-0209467P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Birse CE, Rosen CA;  
XX  
DR WPI; 2002-147878/19.  
XX  
PT N-PSDB; ABQ55674.  
XX  
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
XX useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
PT cancer), immune disorders, cardiovascular disorders and neurological  
PT diseases.  
XX  
PS Claim 11; SEQ ID NO 3729; 2922pp; English.  
XX  
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovary and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents a human ovarian antigen of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 120 AA;  
  
Query Match 81.8%; Score 36; DB 5; Length 120;  
Best Local Similarity 75.0%; Pred. No. 80;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YEDEINNK 8  
Db 37 YEDEINHNR 44  
|||||:  
  
RESULT 9  
AAB87629  
ID AAB87629 standard; protein; 217 AA.  
XX  
AC AAB87629;  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Bovine mammary tissue derived protein #20.

KW Bovine; mammary gland; cancer; tumour; angiogenesis.  
XX  
OS Bos taurus.  
XX  
PN WO200114553-A1.  
XX  
PD 01-MAR-2001.  
XX  
PF 23-AUG-2000; 2000WO-NZ000166.  
XX  
PR 23-AUG-1999; 99US-0150330P.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.  
PI Havukkala IJ, Gleen M, Grigor MR, Molenaar AJ;  
XX WPI; 2001-226619/23.  
DR  
XX  
PT New polypeptides and polynucleotides encoding the polypeptides, which are  
PT expressed in bovine mammary gland tissue, useful for stimulating mammary  
PT gland growth or function, or inducing differentiation of milk producing  
PT cells.  
XX  
PS Claim 11; Page 67-68; 97pp; English.  
XX  
CC The present invention relates to proteins derived from bovine mammary  
CC gland cells. The invention is useful for stimulating bovine mammary  
CC cell growth and function, inhibiting the growth of various mammary gland  
CC cancer cells, inhibiting angiogenesis and vascularization of tumours, or  
CC modulating the growth of blood vessels in a mammal  
XX  
SQ Sequence 217 AA;  
  
Query Match 81.8%; Score 36; DB 4; Length 217;  
Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YEDEINNK 8  
Db 2 YEDEINHNR 9  
|||||:  
  
RESULT 10  
AAG85031  
ID AAG85031 standard; protein; 368 AA.  
XX  
AC AAG85031;  
XX  
DT 06-AUG-2003 (revised)  
DT 11-SEP-2001 (first entry)  
XX  
DE Shrimp white spot Bacilliform virus (WSBV) protein 122.  
XX  
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
KW antiviral agent; gene expression; antisense construct;  
KW transgenic viral resistant shrimp.  
XX  
OS White spot syndrome virus.  
XX  
PN WO200139351-A2.  
XX  
PD 31-MAY-2001.  
XX  
PF 08-NOV-2000; 2000WO-US028888.  
XX  
PR 24-NOV-1999; 99CN-00124717.  
XX  
PA (PENY-) PE CORP NY.  
XX (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
XX (SINO-) SINOGENOMAX CO LTD.  
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;

```

XX WPI; 2001-355877/37.
DR N-FSDB; AAH62811.
XX
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT (WSBV), useful for producing viral polypeptides that can be used to
PT screen for agents that are useful for treating WSBV infection.
XX
XX Claim 1; Fig 3; 626pp; English.
XX
CC The invention provides the primary nucleotide sequence of the WSBV genome
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection, in screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viral gene expression in infected cells and tissues and to create
CC transgenic viral resistant shrimp. (Updated on 06-AUG-2003 to correct OS
CC field.)
XX
SQ Sequence 368 AA;
    Query Match      81.8%; Score 36; DB 4; Length 368;
    Best Local Similarity 75.0%; Pred. No. 2.8e+02;
    Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINNK 8
Db      63 FKDEINNK 70

RESULT 11
ADX94912
ID ADX94912 standard; protein; 370 AA.
XX
AC ADX94912;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polypeptide seqid 57576.
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIU J) LIU J.
PA (ZHOU Y) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
DR

```

```

XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX Claim 1; SEQ ID NO 57576; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 370 AA;
    Query Match      81.8%; Score 36; DB 8; Length 370;
    Best Local Similarity 62.5%; Pred. No. 2.8e+02;
    Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINNK 8
Db      159 FEDEVNKR 166

RESULT 12
ADT57158
ID ADT57158 standard; protein; 407 AA.
XX
AC ADT57158;
XX
DT 13-JAN-2005 (first entry)
XX
DE Plant polypeptide, SEQ ID 7235.
XX
KW Plant; transgenic; cold tolerance; growth rate; drought tolerance;
KW disease resistance; galactomannan production; plant growth regulator;
KW heat tolerance; herbicide tolerance; lignin production;
KW extreme osmotic condition tolerance; pathogens resistance;
KW pest resistance; yield improvement; seed oil yield; seed protein yield.
XX
OS Viridiplantae.
XX
PN US2004216190-A1.
XX
PD 28-OCT-2004.
XX
PF 18-DEC-2003; 2003US-00739930.
XX
PR 28-APR-2003; 2003US-00424599.
PR 28-APR-2003; 2003US-00425115.
XX
PA (KOVA/) KOVALIC D K.
PA Kovalic DK;
XX
DR WPI; 2004-757369/74.
XX
PT New recombinant DNA constructs useful in the field of biochemistry and
PT Genetics, and in particular for producing transgenic plants with improved

```

PT biological characteristics.  
PS Claim 2; SEQ ID NO 7235; 14pp; English.  
XX  
XX  
CC The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomanan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX  
XX  
SQ Sequence 407 AA;  
  
Query Match 81.8%; Score 36; DB 8; Length 407;  
Best Local Similarity 62.5%; Pred. No. 3.1e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YEDEINNK 8  
Db :|||:|:  
196 FEDEVNRR 203  
  
RESULT 13  
ABM81385  
ID ABM81385 standard; protein; 469 AA.  
XX  
XX AC ABM81385;  
XX  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Tumour-associated antigenic target (TAT) polypeptide PRO82152, SEQ:3587.  
XX  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic.  
XX  
XX Homo sapiens.  
OS

PN WO2004030615-A2.  
XX  
PD 15-APR-2004.  
XX  
PF 29-SEP-2003; 2003WO-US028547.  
XX  
PR 02-OCT-2002; 2002US-0414971P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Wu TD, Zhang Z, Zhou Y;  
XX WPI; 2004-347921/32.  
XX N-PSDB; ACN39445.  
DR  
DR New tumor-associated antigenic target polypeptides and nucleic acids,  
XX useful in preparing a medicament for treating or detecting a  
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
XX prostate cancer or tumor.  
XX  
PS Claim 12; SEQ ID NO 3587; 7273pp; English.  
XX  
XX The invention relates to human tumour-associated antigenic target (TAT)  
XX polypeptides, and their related nucleic acids. The TAT polypeptides are  
XX overexpressed in cancer tissues compared to normal tissues, and may thus  
XX serve as effective targets for the diagnosis and treatment of cancer in  
XX mammals. The invention also relates to nucleic acid and polypeptide  
XX sequences at least 80% identical to the TAT nucleic acids and  
XX polypeptides; expression vectors and host cells comprising a TAT nucleic  
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic  
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a  
XX TAT polypeptide; and methods and compositions for the treatment or  
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
XX antibodies, antagonists, binding molecules and compositions are useful  
XX for diagnosing or treating a cell proliferative disorder associated with  
XX increased TAT expression, particularly cancers such as breast cancer,  
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
XX cancer, pancreatic cancer, cervical cancer, cancers of the central  
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
XX used as hybridisation probes, in chromosome and gene mapping, in  
XX chromosome identification and in gene therapy. The present sequence  
XX represents a TAT polypeptide of the invention  
XX  
SQ Sequence 469 AA;  
  
Query Match 81.8%; Score 36; DB 8; Length 469;  
Best Local Similarity 75.0%; Pred. No. 3.6e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YEDEINNK 8  
Db :|||:|:  
180 YEDEINHR 187  
  
RESULT 14  
AAV52397  
ID AAV52397 standard; protein; 546 AA.  
XX  
XX AC AAV52397;  
XX  
XX  
DT 15-FEB-2000 (first entry)  
XX  
XX Human keratin KERT-1.  
DE  
XX Keratin; KERT-1; pancreatic tumour; intermediate filament; consensus;  
KW keratin type II; basic; cytoskeleton; epithelial cell;  
KW differential expression; differentiation; development; proliferation;  
KW wound healing; psoriasis; viral infection; cancer; embryogenesis; skin;  
KW nail; hair; diagnosis; prevention; treatment; epithelial disorder;  
KW developmental disorder; proliferative disorder; reproductive disorder;  
KW antagonist; antibody; detection.  
XX  
XX Homo sapiens.  
OS

XX Key Location/Qualifiers  
 FT Modified-site 8 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 71 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 75 /note= "Phosphorylated by cAMP/cGMP-dependent protein kinase"  
 FT Modified-site 82 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 146 /note= "Glycosylated"  
 FT Modified-site 181 /note= "Phosphorylated by casein kinase II"  
 FT Modified-site 208 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 211 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 245 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 265 /note= "Phosphorylated by casein kinase II"  
 FT Modified-site 282 /note= "Phosphorylated by casein kinase II"  
 FT Modified-site 284 /note= "Phosphorylated by tyrosine kinase"  
 FT Modified-site 304 /note= "Phosphorylated by casein kinase II"  
 FT Modified-site 306 /note= "Phosphorylated by casein kinase II"  
 FT Modified-site 352 /note= "Glycosylated"  
 FT Modified-site 360 /note= "Phosphorylated by casein kinase II"  
 FT Modified-site 362 /note= "Phosphorylated by tyrosine kinase"  
 FT Modified-site 381 /note= "Phosphorylated by casein kinase II"  
 FT Modified-site 457 /note= "Phosphorylated by casein kinase II"  
 FT Region 466. .474 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 468 /note= "Intermediate filament signature motif"  
 FT Modified-site 489 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 494 /note= "Glycosylated"  
 FT Modified-site 533 /note= "Glycosylated"  
 FT Modified-site 540 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site 543 /note= "Phosphorylated by protein kinase C"  
 FT Modified-site /note= "Phosphorylated by protein kinase C or casein kinase II"  
 XX US5994081-A.  
 XX 30-NOV-1999.  
 XX 27-APR-1998; 98US-00067351.  
 XX 27-APR-1998; 98US-00067351.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Tang YT, Hillman JL, Baughn M, Corley NC;  
 XX WPI; 2000-038257/03.  
 XX N-PSDB; AA238337.  
 XX New human keratin polynucleotide and polypeptide useful in diagnosis,

PT prevention and treatment of epithelial, cell proliferative, developmental  
 XX and reproductive disorders.  
 XX Claim 1; Fig 1; 46pp; English.  
 CC This sequence represents a human keratin, KERT-1. Nucleotides encoding  
 CC KERT-1 were initially identified in a pancreatic tumour cDNA library, the  
 CC cDNA encoding this sequence being a consensus. KERT-1 has chemical and  
 CC structural homology (77%) with human keratin type II (basic) #5  
 CC (G307082). Keratins are members of the intermediate filament superfamily  
 CC and are cytoskeletal proteins that assemble into filaments approximately  
 CC 10nm in diameter that give structural support to epithelial cells. About  
 CC thirty keratin genes have so far been identified, and these genes are  
 CC differentially expressed in epithelial tissues at various stages of  
 CC differentiation and development. Keratins are associated with cell  
 CC proliferation, with expression of certain types being induced in  
 CC hyperproliferating stratified epithelia, such as in wound healing,  
 CC psoriasis, viral infection and cancer. They may also have roles in  
 CC development and embryogenesis. Mutations in keratin genes are responsible  
 CC for a variety of skin, nail and/or hair diseases. KERT proteins  
 CC nucleotides can be used for the diagnosis, prevention and treatment of  
 CC epithelial, cell proliferative, developmental and reproductive disorders.  
 CC Epithelial disorders which can be treated include dermatitis, keratosis,  
 CC basal cell carcinoma, herpes simplex, scabies, insect bites,  
 CC dermatofibroma, xerosis, eczema, psoriasis, warts, acne, lupus  
 CC erythematosus, alopecia, skin lesions, and cutaneous drug reactions.  
 CC Developmental disorders which can be treated by KERT include anaemia,  
 CC Cushing's syndrome, myelodysplastic syndrome, cerebral palsy, spina bifida  
 CC or prevent a cell proliferative disorder such as arteriosclerosis,  
 CC atherosclerosis, cirrhosis, hepatitis, myelofibrosis, cancers and  
 CC psoriasis and reproductive disorders including infertility, ovulatory  
 CC defects, endometriosis, cancer of the breast, testis, prostate and  
 CC gynaecomastia. Purified KERT protein can also be used to produce  
 CC antibodies or to screen libraries of pharmaceutical agents to discover  
 CC agents that specifically bind KERT. KERT nucleotides can be used in  
 CC therapeutic compositions e.g., for gene therapy  
 XX Sequence 546 AA;  
 SQ  
 Query Match 81.8%; Score 36; DB 3; Length 546;  
 Best Local Similarity 75.0%; Pred. No. 4.3e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 YEDEINNK 8  
 Db 257 YEDEINHR 264  
 RESULT 15  
 AAEE20422  
 ID AAEE20422 standard; protein; 546 AA.  
 XX AC AAEE20422;  
 XX 01-JUL-2002 (first entry)  
 XX Human keratin-1 (KER-1) protein.  
 XX Human, keratin-1; KERT-1; vitiligo; epithelial disorder; developmental;  
 KW reproductive; cell proliferative; atopic dermatitis; epidermolysis;  
 KW eczema; arteriosclerosis; cirrhosis; hepatitis; cerebral palsy; glaucoma;  
 KW muscular dystrophy; cataract; infertility; ovarian tumour; testis cancer;  
 KW ectopic pregnancy; prostatitis; drug screening.  
 XX Homo sapiens.  
 XX US6221843-B1.  
 XX 24-APR-2001.  
 XX 23-JUL-1999; 99US-00360490.  
 XX

PR 27-APR-1998; 98US-00067351.  
XX  
XX PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX PI Tang YT, Hillman JL, Corley NC, Baughn M;  
XX  
XX DR WPI; 2002-279090/32.  
XX N-PSDB; AAD32706.  
XX  
XX Novel human keratin protein designated KERT1-3, useful e.g. for  
PT diagnosing, treating and preventing vitiligo, eczema, cirrhosis,  
PT hepatitis, cerebral palsy, congenital glaucoma, cataract, prostatitis and  
PT infertility.  
XX  
XX Claim 1; Fig 1; 46pp; English.  
XX  
XX The invention relates to a substantially purified human keratin (KERT)  
CC polypeptide designated KERT1-3. KERT protein is useful for screening a  
CC library of molecules and compounds to identify a molecule or compound  
CC which specifically binds to the protein. The method involves contacting  
CC KERT polypeptide with a library of molecules or compounds e.g., a library  
CC of peptides or proteins, agonists, antagonists, antibodies, inhibitors,  
CC immunoglobulins, drug compounds, and pharmaceutical agents, under  
CC conditions that allow specific binding; and detecting specific binding,  
CC thereby identifying a molecule or compound which specifically binds the  
CC protein. KERT protein is useful for diagnosis, treatment and prevention  
CC of epithelial disorders (vitiligo, atopic dermatitis, epidermolysis  
CC bullosa simplex, eczema); cell proliferative disorder (arteriosclerosis,  
CC cirrhosis, hepatitis); developmental disorder (Cerebral palsy, Duchenne  
CC and Becker muscular dystrophy, myelodysplastic syndrome, congenital  
CC glaucoma, cataract) and reproductive disorder (infertility, endometrial  
CC and ovarian tumours, ectopic pregnancies, cancer of the testis, benign  
CC prostatic hyperplasia, prostatitis). KERT is also useful for screening  
CC libraries or libraries of compounds in drug screening techniques. The  
CC present sequence is human KERT-1 protein  
XX  
SQ Sequence 546 AA;

Query Match 81.8%; Score 36; DB 5; Length 546;  
Best Local Similarity 75.0%; Pred. NO. 4.3e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
DB 257 YEDEINHR 264

Search completed: July 5, 2006, 18:43:43  
Job time : 72.5146 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 58.5631 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-11  
Perfect score: 44  
Sequence: 1 YEDEINNK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	8	5	US-10-774-928-11
2	38	86.4	270	4	US-10-425-115-190930
3	38	86.4	645	4	US-10-425-115-280884
4	37	84.1	112	4	US-10-767-701-37605
5	37	84.1	246	4	US-10-425-115-315519
6	37	84.1	295	4	US-10-767-701-41541
7	37	84.1	526	4	US-10-425-115-295632
8	37	84.1	612	4	US-10-437-963-103055
9	37	84.1	703	5	US-10-739-930-10003
10	36	81.8	120	4	US-10-264-049-3729
11	36	81.8	217	4	US-10-263-828-88
12	36	81.8	352	5	US-10-617-316-244
13	36	81.8	370	4	US-10-425-114-57576
14	36	81.8	407	5	US-10-739-930-7235
15	36	81.8	488	4	US-10-425-115-268674
16	35	79.5	113	5	US-10-480-848A-5
17	35	79.5	139	3	US-09-351-794A-4
18	35	79.5	186	3	US-09-864-761-36050
19	35	79.5	186	5	US-10-450-763-39769
20	35	79.5	222	4	US-10-424-599-263693
21	35	79.5	236	4	US-10-264-049-4279
22	35	79.5	273	5	US-10-450-763-56735
23	35	79.5	275	5	US-10-450-763-45641
24	35	79.5	398	5	US-10-883-020-5
25	35	79.5	422	3	US-09-779-307-18
26	35	79.5	422	3	US-09-779-307-20
27	35	79.5	429	5	US-10-450-763-49905

28 35 79.5 429 5 US-10-450-763-49906 Sequence 49906, A  
29 35 79.5 430 3 US-09-779-307-19 Sequence 19, Appl  
30 35 79.5 433 3 US-09-925-298-463 Sequence 463, App  
31 35 79.5 433 4 US-10-102-806-463 Sequence 463, App  
32 35 79.5 469 4 US-10-093-766-59 Sequence 59, Appl  
33 35 79.5 469 4 US-10-097-340-171 Sequence 171, App  
34 35 79.5 469 4 US-10-097-340-173 Sequence 173, App  
35 35 79.5 469 5 US-10-733-969A-52 Sequence 52, Appl  
36 35 79.5 469 6 US-11-050-926-171 Sequence 171, App  
37 35 79.5 469 6 US-11-050-926-173 Sequence 173, App  
38 35 79.5 476 3 US-09-779-307-16 Sequence 16, Appl  
39 35 79.5 476 5 US-10-821-234-1064 Sequence 1064, Ap  
40 35 79.5 477 3 US-09-779-307-4 Sequence 4, Appl  
41 35 79.5 481 4 US-10-080-334-54 Sequence 54, Appl  
42 35 79.5 481 6 US-11-156-300-24 Sequence 24, Appl  
43 35 79.5 482 3 US-09-779-307-17 Sequence 17, Appl  
44 35 79.5 482 4 US-10-080-334-187 Sequence 187, App  
45 35 79.5 482 5 US-10-883-020-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-10-774-928-11  
; Sequence 11, Application US/10774928  
; Publication NO. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-11

Query Match 100.0%; Score 44; DB 5; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.9e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8  
Db 1 YEDEINNK 8

RESULT 2  
US-10-425-115-190930  
; Sequence 190930, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Placens  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 190930  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(270)

```
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_105706C.1.pep
US-10-425-115-190930

Query Match      86.4%; Score 38; DB 4; Length 270;
Best Local Similarity 87.5%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8
Db 88 YEDEINEK 95
|||||

RESULT 3
US-10-425-115-280884
; Sequence 280884, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 280884
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(246)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_19268C.1.pep
US-10-425-115-280884

Query Match      86.4%; Score 38; DB 4; Length 645;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8
Db 185 YEDEINEK 192
|||||

RESULT 4
US-10-767-701-37605
; Sequence 37605, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 37605
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C3323_1.pep
US-10-767-701-37605

Query Match      84.1%; Score 37; DB 4; Length 112;
Best Local Similarity 85.7%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7
Db 194 YEDELNN 200
|||||

RESULT 5
US-10-425-115-315519
; Sequence 315519, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 315519
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(246)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_50820C.1.pep
US-10-425-115-315519

Query Match      84.1%; Score 37; DB 4; Length 246;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7
Db 189 YEDELNN 195
|||||

RESULT 6
US-10-767-701-41541
; Sequence 41541, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 41541
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C446_1.pep
US-10-767-701-41541

Query Match      84.1%; Score 37; DB 4; Length 295;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7
Db 194 YEDELNN 200
|||||

RESULT 7
US-10-425-115-295632
```



; Sequence 295632, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 295632  
; LENGTH: 526  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(526)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_32695C.1.p  
US-10-425-115-295632

Query Match 84.1%; Score 37; DB 4; Length 526;  
Best Local Similarity 85.7%; Pred. No. 2.8e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
|||:|  
Db 424 YEDELNN 430

RESULT 8  
US-10-437-963-103055  
; Sequence 103055, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 103055  
; LENGTH: 612  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(612)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_100522C.1.p  
US-10-437-963-103055

Query Match 84.1%; Score 37; DB 4; Length 612;  
Best Local Similarity 85.7%; Pred. No. 3.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
|||:|  
Db 510 YEDELNN 516

## RESULT 9

US-10-739-930-10003  
; Sequence 10003, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 10003  
; LENGTH: 703  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(703)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-C236\_1.p  
US-10-739-930-10003

Query Match 84.1%; Score 37; DB 5; Length 703;  
Best Local Similarity 85.7%; Pred. No. 3.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
|||:|  
Db 250 YEDELNN 256

## RESULT 10

US-10-264-049-3729  
; Sequence 3729, Application US/10264049  
; Publication No. US20040005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA133P1  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: US 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 4360  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 3729  
; LENGTH: 120  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (65)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
US-10-264-049-3729

Query Match 81.8%; Score 36; DB 4; Length 120;  
Best Local Similarity 75.0%; Pred. No. 89;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 8  
|||:|  
Db 37 YEDEINHR 44

```
RESULT 11
US-10-823-828-88
; Sequence 88, Application US/10263828
; Publication No. US20030138905A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: mammary gland and methods for their use.
; FILE REFERENCE: 11000.1044U1con
; CURRENT APPLICATION NUMBER: US/10/263,828
; CURRENT FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 88
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Bovine
US-10-263-828-88

Query Match      81.8%; Score 36; DB 4; Length 217;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINNK 8
Db      2 YEDEINHR 9

RESULT 12
US-10-617-316-244
; Sequence 244, Application US/10617316
; Publication No. US20050130263A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions Isolated From Bovine
; TITLE OF INVENTION: Mammary Gland and Methods For Their Use.
; FILE REFERENCE: 11000.1046U1c1
; CURRENT APPLICATION NUMBER: US/10/617,316
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 09/699,146
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/162,701
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 262
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Bovine
US-10-617-316-244

Query Match      81.8%; Score 36; DB 5; Length 352;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINNK 8
Db      67 YEDEINHR 74

RESULT 13
US-10-425-114-57576
; Sequence 57576, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57576
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMROTEOSINT026E11_FLI.ppep
US-10-425-114-57576

Query Match      81.8%; Score 36; DB 4; Length 370;
Best Local Similarity 62.5%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINNK 8
Db      159 FEDEVNHR 166

RESULT 14
US-10-739-930-7235
; Sequence 7235, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 7235
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C10961_25.p
US-10-739-930-7235

Query Match      81.8%; Score 36; DB 5; Length 407;
Best Local Similarity 62.5%; Pred. No. 3.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy      1 YEDEINNK 8
Db      196 FEDEVNHR 203

RESULT 15
US-10-425-115-268674
; Sequence 268674, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 268674
; LENGTH: 486
```

```

; TYPE: PRT
;
; ORGANISM: Zea mays
;
; FEATURE:
;
; OTHER INFORMATION: Clone ID: MRT4577_17662C.1.pep
US-10-425-115-268674

```

```
Query Match      81.8%; Score 36; DB 4; Length 488;
Best Local Similarity 62.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels
```

Qy	1 YEDEINNK 8
	:   : :
Db	277 FEDEVNRR 28

Search completed: July 5, 2006, 20:51:55  
Job time : 59.5631 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 4.58252 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-11  
Perfect score: 44  
Sequence: 1 YEDEINNK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New.\*  
1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	84.1	714	6	US-10-449-902-54951 Sequence 54951, A
2	35	79.5	469	6	US-10-505-928-457 Sequence 457, App
3	35	79.5	483	7	US-11-105-233-180 Sequence 180, App
4	34	77.3	758	6	US-10-449-902-51086 Sequence 51086, A
5	34	77.3	799	6	US-10-449-902-47602 Sequence 47602, A
6	34	77.3	1550	6	US-10-471-571A-1300 Sequence 1300, App
7	33	75.0	33	7	US-11-134-871-915 Sequence 915, App
8	33	75.0	940	6	US-10-449-902-45238 Sequence 45238, A
9	33	75.0	940	6	US-10-449-902-45669 Sequence 45669, A
10	32	72.7	26	7	US-11-134-871-1603 Sequence 1603, App
11	32	72.7	420	6	US-10-953-349-23057 Sequence 23057, A
12	32	72.7	430	6	US-10-953-349-23056 Sequence 23056, A
13	32	72.7	533	6	US-10-449-902-33245 Sequence 33245, A
14	32	72.7	533	6	US-10-449-902-43690 Sequence 43690, A
15	32	72.7	599	6	US-10-449-902-49534 Sequence 49534, A
16	32	72.7	964	7	US-11-289-102-392 Sequence 392, App
17	31	70.5	247	6	US-10-449-902-54216 Sequence 54216, A
18	31	70.5	568	6	US-10-449-902-46926 Sequence 46926, A
19	31	70.5	568	6	US-10-449-902-47019 Sequence 47019, A
20	31	70.5	574	6	US-10-449-902-12941 Sequence 12941, A
21	31	70.5	636	6	US-10-953-349-12941 Sequence 12941, A
22	31	70.5	636	6	US-10-953-349-12940 Sequence 12940, A
23	30	68.2	91	6	US-10-953-349-36767 Sequence 36767, A
24	30	68.2	107	6	US-10-953-349-13561 Sequence 13561, A
25	30	68.2	132	6	US-10-953-349-36766 Sequence 36766, A

26	30	68.2	140	6	US-10-953-349-1458 Sequence 1458, App
27	30	68.2	142	6	US-10-953-349-24110 Sequence 24110, A
28	30	68.2	147	6	US-10-953-349-1457 Sequence 1457, App
29	30	68.2	148	6	US-10-953-349-36765 Sequence 36765, A
30	30	68.2	158	6	US-10-953-349-13251 Sequence 13251, A
31	30	68.2	158	6	US-10-953-349-13560 Sequence 13560, A
32	30	68.2	164	6	US-10-953-349-1456 Sequence 1456, App
33	30	68.2	183	6	US-10-953-349-13250 Sequence 13250, A
34	30	68.2	201	7	US-11-199-489A-68 Sequence 68, Appl
35	30	68.2	228	6	US-10-449-902-50750 Sequence 50750, A
36	30	68.2	245	6	US-10-953-349-24109 Sequence 24109, A
37	30	68.2	288	6	US-10-953-349-22862 Sequence 22862, A
38	30	68.2	295	6	US-10-449-902-48257 Sequence 48257, A
39	30	68.2	350	6	US-10-449-902-55155 Sequence 55155, A
40	30	68.2	358	6	US-10-953-349-22861 Sequence 22861, A
41	30	68.2	412	6	US-10-471-571A-84 Sequence 84, Appl
42	30	68.2	434	6	US-10-953-349-22860 Sequence 22860, A
43	30	68.2	450	6	US-10-449-902-36104 Sequence 36104, A
44	30	68.2	472	6	US-10-196-749-98 Sequence 98, Appl
45	30	68.2	472	7	US-11-190-075-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-10-449-902-54951  
; Sequence 54951, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205V1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 54951  
; LENGTH: 714  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-54951

Query Match 84.1%; Score 37; DB 6; Length 714;  
Best Local Similarity 85.7%; Pred. No. 18;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
DB 613 YEDELNN 619

RESULT 2  
US-10-505-928-457  
; Sequence 457, Application US/10505928  
; Publication No. US2006008832A1  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig Institute for Cancer Research et al.  
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES  
; FILE REFERENCE: 28967/39178  
; CURRENT APPLICATION NUMBER: US/10/505,928  
; CURRENT FILING DATE: 2004-08-27  
; PRIOR APPLICATION NUMBER: US 60/363,019  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: PatentIn 3.2

; SEQ ID NO 457  
; LENGTH: 469  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-505-928-457

Query Match 79.5%; Score 35; DB 6; Length 469;  
Best Local Similarity 75.0%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8  
Db 180 YEDEINRR 187

## RESULT 3

US-11-105-233-180  
; Sequence 180, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Thiagalingam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; FILE REFERENCE: Tumors  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; PRIOR FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-105-233-180

Query Match 79.5%; Score 35; DB 7; Length 483;  
Best Local Similarity 75.0%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YEDEINNK 8  
Db 179 YEDEINKR 186

## RESULT 4

US-10-449-902-51086  
; Sequence 51086, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; PRIOR FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51086  
; LENGTH: 758  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-51086

Query Match 77.3%; Score 34; DB 6; Length 758;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7

Db 230 YEDEIND 236

## RESULT 5

US-10-449-902-47602  
; Sequence 47602, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; PRIOR FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 47602  
; LENGTH: 799  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-47602

Query Match 77.3%; Score 34; DB 6; Length 799;  
Best Local Similarity 85.7%; Pred. No. 73;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7  
Db 349 YEDEIND 355

## RESULT 6

US-10-471-571A-1300  
; Sequence 1300, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927WO  
; CURRENT APPLICATION NUMBER: US/10/471,571A  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 1300  
; LENGTH: 1550  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1)..(1550)  
; OTHER INFORMATION: phi PVL ORF 15 and 16  
US-10-471-571A-1300

Query Match 77.3%; Score 34; DB 6; Length 1550;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEDEINN 7  
Db 1522 FEDEVN 1528

## RESULT 7

US-11-134-871-915

; Sequence 915, Application US/11134871  
; Publication No. US20060141528A1  
; GENERAL INFORMATION:  
; APPLICANT: Aebersold, Rudolf H.  
; APPLICANT: Zhang, Hui  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Quatification of Serum Glycoproteins  
; FILE REFERENCE: 66661-116  
; CURRENT APPLICATION NUMBER: US/11/134, 871  
; CURRENT FILING DATE: 2005-05-20  
; PRIOR APPLICATION NUMBER: 60/573,593  
; PRIOR FILING DATE: 2004-05-21  
; NUMBER OF SEQ ID NOS: 3602  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 915  
; LENGTH: 940  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-134-871-915

Query Match 75.0%; Score 33; DB 7; Length 940;  
Best Local Similarity 71.4%; Pred. No. 3.6;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
Db 20 YEDEVNS 26

## RESULT 8

US-10-449-902-45238  
; Sequence 45238, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449, 902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 45238  
; LENGTH: 940  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-45238

Query Match 75.0%; Score 33; DB 6; Length 940;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEIN 6  
Db 413 YEDEIN 418

## RESULT 9

US-10-449-902-45669  
; Sequence 45669, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF

; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 45669  
; LENGTH: 940  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-45669

Query Match 75.0%; Score 33; DB 6; Length 940;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEIN 6  
Db 413 YEDEIN 418

## RESULT 10

US-11-134-871-1603  
; Sequence 1603, Application US/11134871  
; Publication No. US20060141528A1  
; GENERAL INFORMATION:  
; APPLICANT: Aebersold, Rudolf H.  
; APPLICANT: Zhang, Hui  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Quatification of Serum Glycoproteins  
; FILE REFERENCE: 66661-116  
; CURRENT APPLICATION NUMBER: US/11/134,871  
; CURRENT FILING DATE: 2005-05-20  
; PRIOR APPLICATION NUMBER: 60/573,593  
; PRIOR FILING DATE: 2004-05-21  
; NUMBER OF SEQ ID NOS: 3602  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1603  
; LENGTH: 26  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-134-871-1603

Query Match 72.7%; Score 32; DB 7; Length 26;  
Best Local Similarity 85.7%; Pred. No. 4.3;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
Db 3 YEDEPNN 9

## RESULT 11

US-10-953-349-23057  
; Sequence 23057, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 23057  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-953-349-23057

```
Query Match          72.7%; Score 32; DB 6; Length 420;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EDEINN 7
   |||||
Db 31 EDEINN 36

RESULT 12
US-10-953-349-23056
; Sequence 23056, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 23056
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-23056

Query Match          72.7%; Score 32; DB 6; Length 430;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EDEINN 7
   |||||
Db 41 EDEINN 46

RESULT 13
US-10-449-902-33245
; Sequence 33245, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33245
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-33245

Query Match          72.7%; Score 32; DB 6; Length 533;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DEINN 8
   |||||
Db 444 DEINN 449

RESULT 14
US-10-449-902-33245
; Sequence 33245, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33245
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-33245

Query Match          72.7%; Score 32; DB 6; Length 533;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DEINN 8
   |||||
Db 444 DEINN 449
```

```
US-10-449-902-43690
; Sequence 43690, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43690
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-43690

Query Match          72.7%; Score 32; DB 6; Length 533;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DEINN 8
   |||||
Db 444 DEINN 449

RESULT 15
US-10-449-902-49534
; Sequence 49534, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49534
; LENGTH: 599
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-49534

Query Match          72.7%; Score 32; DB 6; Length 599;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7
   :|||||
Db 134 FQDEINN 140

Search completed: July 5, 2006, 20:53:00
Job time : 5.58252 secs
```



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 18.0194 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-11

Perfect score: 44  
Sequence: 1 YEDEINNK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H\_COMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE\_COMB.pep.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	8	2 US-09-809-003A-11	Sequence 11, Appl
2	36	81.8	546	1 US-09-067-351-1	Sequence 1, Appli
3	36	81.8	546	2 US-09-360-490-1	Sequence 1, Appli
4	35	79.5	139	2 US-08-559-896B-4	Sequence 4, Appli
5	35	79.5	139	2 US-09-351-794A-4	Sequence 4, Appli
6	35	79.5	384	2 US-09-949-016-11034	Sequence 11034, A
7	35	79.5	469	2 US-09-077-606-3	Sequence 3, Appli
8	35	79.5	476	2 US-09-949-016-9096	Sequence 9096, Ap
9	35	79.5	482	2 US-09-538-092-858	Sequence 858, App
10	35	79.5	483	2 US-09-919-497-79	Sequence 79, Appl
11	35	79.5	502	2 US-09-949-016-11033	Sequence 11033, A
12	35	79.5	551	1 US-09-067-351-2	Sequence 2, Appli
13	35	79.5	551	2 US-09-360-490-2	Sequence 2, Appli
14	35	79.5	554	2 US-09-949-016-6628	Sequence 6628, Ap
15	35	79.5	567	2 US-10-104-047-3772	Sequence 3772, Ap
16	35	79.5	569	2 US-09-949-016-11035	Sequence 11035, A
17	35	79.5	569	2 US-09-949-016-11036	Sequence 11036, A
18	35	79.5	637	2 US-09-949-016-8152	Sequence 8152, Ap
19	35	79.5	643	2 US-09-538-092-844	Sequence 844, App
20	35	79.5	645	2 US-09-919-172-41	Sequence 41, Appl
21	35	79.5	652	2 US-08-559-896B-2	Sequence 2, Appli
22	35	79.5	652	2 US-09-351-794A-2	Sequence 2, Appli
23	34	77.3	67	2 US-09-248-796A-23567	Sequence 23567, A
24	34	77.3	220	2 US-09-134-000C-4085	Sequence 4085, Ap
25	33	75.0	7	2 US-09-809-003A-1	Sequence 1, Appli
26	33	75.0	8	2 US-09-809-003A-12	Sequence 12, Appl

27 33 75.0 389 2 US-09-495-406-20 Sequence 20, Appl  
28 33 75.0 389 2 US-09-816-028A-34 Sequence 34, Appl  
29 33 75.0 389 2 US-10-303-162-34 Sequence 34, Appl  
30 33 75.0 389 2 US-10-303-134-34 Sequence 34, Appl  
31 33 75.0 389 2 US-10-303-118-34 Sequence 34, Appl  
32 33 75.0 389 2 US-10-303-128-34 Sequence 34, Appl  
33 33 75.0 389 3 US-10-735-419-34 Sequence 34, Appl  
34 33 75.0 443 2 US-09-949-016-10582 Sequence 10582, A  
35 33 75.0 631 1 US-08-487-890A-115 Sequence 115, App  
36 33 75.0 631 1 US-08-478-435-115 Sequence 115, App  
37 33 75.0 631 1 US-08-337-483-115 Sequence 115, App  
38 33 75.0 631 1 US-08-478-373-115 Sequence 115, App  
39 33 75.0 631 2 US-08-474-671-115 Sequence 115, App  
40 33 75.0 631 2 US-08-483-577A-115 Sequence 115, App  
41 33 75.0 631 2 US-08-897-438-115 Sequence 115, App  
42 33 75.0 631 2 US-08-637-654-115 Sequence 115, App  
43 33 75.0 631 2 US-08-649-518-115 Sequence 115, App  
44 33 75.0 995 2 US-09-362-842-2 Sequence 2, Appli  
45 33 75.0 1348 2 US-09-949-002-517 Sequence 517, App

## ALIGNMENTS

## RESULT 1

US-09-809-003A-11  
; Sequence 11, Application US/09809003A  
; Patent No. 6673351  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE REFERENCE: Psoriasis  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-09-809-003A-11

Query Match 100.0%; Score 44; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 5e+03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
Db 1 YEDEINNK 8

## RESULT 2

US-09-067-351-1  
; Sequence 1, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/067.351
;; FILING DATE: Herewith
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: CERRONE, MICHAEL C.
;; REGISTRATION NUMBER: 39,132
;; REFERENCE/DOCKET NUMBER: PF-0511 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (650) 855-0555
;; TELEFAX: (650) 845-4166
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 546 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: PANCYTUT02
;; CLONE: 1467090
US-09-067-351-1
```

```
Query Match 81.8%; Score 36; DB 1; Length 546;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 YEDEINNK 8
Db 257 YEDEINHR 264
```

```
RESULT 3
US-09-360-490-1
; Sequence 1, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/360.490
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/067.351
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
```

```
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: PANCYTUT02
;; CLONE: 1467090
US-09-360-490-1
```

```
Query Match 81.8%; Score 36; DB 2; Length 546;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 YEDEINNK 8
Db 257 YEDEINHR 264
```

```
RESULT 4
US-08-559-896B-4
; Sequence 4, Application US/08559896B
; Patent No. 6310046
; GENERAL INFORMATION:
; APPLICANT: Patrick E. Duffy
; APPLICANT: Christian F. Ockenhouse
; TITLE OF INVENTION: SEQUESTIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John Moran
; STREET: USA MPMC - MCMR-JA
; CITY: FORT DETRICK, FREDERICK
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21702-5012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/559.896B
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Moran, John
; REGISTRATION NUMBER: 26,313
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 619-2065
; TELEFAX: (301) 619-7714
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; TOPOLOGY: Linear
US-08-559-896B-4
```

```
Query Match 79.5%; Score 35; DB 2; Length 139;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 YEDEINN 7
Db 84 YKDEINN 90
```

```
RESULT 5
US-09-351-794A-4
; Sequence 4, Application US/09351794A
; Patent No. 6641815
; GENERAL INFORMATION:
```

; APPLICANT: DUFFY, PATRICK E.  
; APPLICANT: OCKENHOUSE, CHRISTIAN F.  
; TITLE OF INVENTION: SEQUESTIN  
; FILE REFERENCE: 38644-175519  
; CURRENT APPLICATION NUMBER: US/09/351,794A  
; CURRENT FILING DATE: 1999-07-13  
; PRIOR APPLICATION NUMBER: 08/559,896  
; PRIOR FILING DATE: 1995-11-17  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 139  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-09-351-794A-4

Query Match 79.5%; Score 35; DB 2; Length 139;  
Best Local Similarity 85.7%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEDEINN 7  
|:|||||  
DB 84 YKDEINN 90

## RESULT 6

US-09-949-016-11034  
; Sequence 11034, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11034  
; LENGTH: 384  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-11034

Query Match 79.5%; Score 35; DB 2; Length 384;  
Best Local Similarity 75.0%; Pred. No. 80;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINN 8  
|:|||||  
DB 276 YEDEINKR 283

## RESULT 7

US-09-077-606-3  
; Sequence 3, Application US/09077606  
; Patent No. 6774220  
; GENERAL INFORMATION:  
; APPLICANT: JIANG, Pan Hong  
; APPLICANT: KABA, Aboubacar  
; APPLICANT: CHANY-FOURNIER, Francoise  
; APPLICANT: CERUTTI, Itallia  
; APPLICANT: CHANY, Charles  
; TITLE OF INVENTION: COMPOUNDS HAVING LECTINIC PROPERTIES AND THEIR  
; FILE REFERENCE: 040388/0113  
; CURRENT APPLICATION NUMBER: US/09/077,606

; CURRENT FILING DATE: 1998-07-30  
; EARLIER APPLICATION NUMBER: WO PCT/FR96/01937  
; EARLIER FILING DATE: 1996-12-04  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 469  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-077-606-3

Query Match 79.5%; Score 35; DB 2; Length 469;  
Best Local Similarity 75.0%; Pred. No. 99;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINN 8  
|:|||||  
DB 180 YEDEINRR 187

## RESULT 8

US-09-949-016-9096  
; Sequence 9096, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9096  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-9096

Query Match 79.5%; Score 35; DB 2; Length 476;  
Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINN 8  
|:|||||  
DB 187 YEDEINRR 194

## RESULT 9

US-09-538-092-858  
; Sequence 858, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratSeqFormatter Version 0.9  
; SEQ ID NO 858  
; LENGTH: 482

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P05787
US-09-538-092-858

Query Match      79.5%; Score 35; DB 2; Length 482;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 0; Gaps 0; Indels 1;

Qy      1 YEDEINNK 8
Db      178 YEDEINKR 185

RESULT 10
US-09-919-497-79
; Sequence 79, Application US/09919497
; Patent No. 6773883
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 79
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-79

Query Match      79.5%; Score 35; DB 2; Length 483;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 0; Gaps 0; Indels 1;

Qy      1 YEDEINNK 8
Db      179 YEDEINKR 186

RESULT 11
US-09-949-016-11033
; Sequence 11033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11033
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11033

Query Match      79.5%; Score 35; DB 2; Length 502;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Gaps 0; Indels 1;

Qy      1 YEDEINNK 8
Db      239 YEDEINKR 246

RESULT 13
US-09-360-490-2
; Sequence 2, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,351
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
; CLONE: 2029060
US-09-067-351-2

Query Match      79.5%; Score 35; DB 1; Length 551;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Gaps 0; Indels 1;

Qy      1 YEDEINNK 8
Db      239 YEDEINKR 246

RESULT 13
US-09-360-490-2
; Sequence 2, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/360,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/067,351  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-360-490-2

Query Match 79.5%; Score 35; DB 2; Length 551;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
DB 235 YEDEINKR 246

RESULT 14  
US-09-949-016-6628  
Sequence 6628, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6628  
LENGTH: 564  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-6628

Query Match 79.5%; Score 35; DB 2; Length 564;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
DB 253 YEDEINKR 260

RESULT 15  
US-10-104-047-3772  
Sequence 3772, Application US/10104047  
Patent No. 6943241  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. 6943241el full length cdna  
FILE REFERENCE: H1-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 3772  
LENGTH: 567  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-104-047-3772

Query Match 79.5%; Score 35; DB 2; Length 567;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YEDEINNK 8  
DB 235 YEDEINKR 242

Search completed: July 5, 2006, 19:24:22  
Job time : 18.0194 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 9.70874 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928A-12  
Perfect score: 43  
Sequence: 1 KYEDEINK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: Pirl:\*  
2: Pirl:\*  
3: Pirl:\*  
4: Pirl:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	310	2 S43865	cyto keratin 8, typ
2	43	100.0	384	2 I61769	keratin 6d, type I
3	43	100.0	390	2 A60093	cyto keratin, type
4	43	100.0	419	1 KRXL2B	keratin, 64K type
5	43	100.0	425	1 KRXL2A	keratin, 64K type
6	43	100.0	461	2 A43782	keratin, type II -
7	43	100.0	483	2 A34720	keratin 8, type II
8	43	100.0	487	2 JT0407	keratin 8, type II
9	43	100.0	490	2 JS0658	cyto keratin EndoA
10	43	100.0	513	2 S08381	keratin, 58K type
11	43	100.0	520	2 JS0291	intermediate filam
12	43	100.0	553	2 IRR009	epidermal keratin
13	43	100.0	564	1 KRHUEA	keratin 6a, type I
14	43	100.0	564	1 KRHUEB	keratin 6b, type I
15	43	100.0	564	2 I61770	keratin 6e, type I
16	43	100.0	564	2 I61768	keratin 6c, type I
17	43	100.0	564	2 I61771	keratin 6f, type I
18	43	100.0	581	1 KRMS2	keratin, type II c
19	43	100.0	590	2 A29904	keratin 5, type II
20	43	100.0	629	2 A29666	keratin, 65K type
21	43	100.0	629	2 S42629	keratin K3 - rabbi
22	43	100.0	638	2 I53169	cyto keratin 2, hu
23	43	100.0	643	1 KRH2	keratin 1, type II
24	43	100.0	645	2 A44861	keratin, 67K type
25	40	93.0	469	2 S05602	keratin K7, type I
26	40	93.0	489	2 B24177	keratin, 55K type
27	40	93.0	502	2 A23547	keratin, type II c
28	39	90.7	164	2 C84777	cyclophilin-like p
29	39	90.7	524	2 A23518	keratin, 57K type

ALIGNMENTS

RESULT 1

S43865

cyto keratin 8, type II - potoro (fragment)

C;Species: Potorous tridactylus (potoroo, long-nosed rat kangaroo)

C;Date: 20-Feb-1995 #sequence\_revision 06-Dec-1996 #text\_change 09-Jul-2004

C;Accession: S43865; S40172

R;Boettger, V.; Lane, E.B.

J. Mol. Biol. 235, 61-67, 1994

A;Title: A monoclonal antibody epitope on keratin 8 identified using a phage peptide lib

A;Reference number: S43865; MUID:94118290; PMID:7507169

A;Accession: S43865

A;Molecule type: mRNA

A;Residues: 1-310 <BOE>

A;Cross-references: UNIPROT:Q28810; UNIPARC:UPI0000086BF0; EMBL:X70987; NID:9437902; PIDN

A;Experimental source: kidney epithelium cells

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil

Query Match 100.0%; Score 43; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8

DB 179 KYEDEINK 186

RESULT 2

I61769

keratin 6d, type II - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C;Accession: I61769

R;Takahashi, K.; Paladini, R.D.; Coulombe, P.A.

J. Biol. Chem. 270, 18581-18592, 1995

A;Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A;Reference number: A57398; MUID:95355491; PMID:7543104

A;Accession: I61769

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-384 <RES>

A;Cross-references: UNIPROT:P48687; UNIPARC:UPI000012DB17; GB:L42602; NID:9914823; GB:L42

28; GB:L42608; NID:9914829; GB:L42609; NID:9914830; GB:L42610; NID:91488952; PIDN:AAB606

C;Genetics:

A;Gene: KRT6D

A;Introns: 72/2; 92/3; 124/3; 179/3; 221/3; 295/2; 307/1

C;Superfamily: cytoskeletal keratin

C;Keywords: coiled coil

Query Match 100.0%; Score 43; DB 2; Length 384;

Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 KYEDEINK 8
Db      72 KYEDEINK 79

RESULT 3
A60093
cytokeratin, type II, early ectodermal - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 30-Sep-1993
C:Accession: A60093
R:Charlebois, T.S.; Spencer, D.H.; Tarkington, S.K.; Henry, J.J.; Grainger, R.M.
Development 108, 33-45, 1990
A:Title: Isolation of a chick cytokeratin cDNA clone indicative of regional specialization
A:Reference number: A60093; MUID:90276238; PMID:1693557
A:Accession: A60093
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-390 <CHA>
A:Cross-references: UNIPARC:UPI00001774C6
C:Superfamily: cytoskeletal keratin
C:Keywords: intermediate filament

Query Match      100.0%; Score 43; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.3; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY      1 KYEDEINK 8
Db      99 KYEDEINK 106

RESULT 4
KXKL2B
keratin, 64K type II cytoskeletal (clone pUF164) - African clawed frog (fragment)
N:Alternate names: 64-kDa type II keratin
C:Species: Xenopus laevis (African clawed frog)
C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 09-Jul-2004
C:Accession: A02952
R:Hoffmann, W.; Franz, J.K.; Franke, W.W.
J. Mol. Biol. 184, 713-724, 1985
A:Title: Amino acid sequence microheterogeneities of basic (type II) cytokeratins of Xenopus laevis
A:Reference number: A2914; MUID:86011576; PMID:2413219
A:Accession: A02952
A:Molecule type: mRNA
A:Residues: 1-419 <HOF>
A:Cross-references: UNIPROT:P04265; UNIPARC:UPI000012DB07; GB:X02895; GB:M13954; NID:964
A:Experimental source: clone pUF164
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-275/Domain: rod (fragment) <ROD>
F:1-9/Region: linker 1
F:10-110/Region: coil 1B
F:111-127/Region: linker 12
F:128-146/Region: coil 2A
F:147-154/Region: linker 2
F:155-275/Region: coil 2B
F:213/Region: stutter
F:276-419/Domain: tail <END>
F:276-395/Region: H2 and V2 subdomains
F:400-419/Region: E2 subdomain

Query Match      100.0%; Score 43; DB 1; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.4; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY      1 KYEDEINK 8
Db      51 KYEDEINK 58

RESULT 5
KXKL2A
keratin, 64K type II cytoskeletal (clone pUF23) - African clawed frog (fragment)
N:Alternate names: 64-kDa type II keratin
C:Species: Xenopus laevis (African clawed frog)
C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 09-Jul-2004
C:Accession: A02953
R:Hoffmann, W.; Franz, J.K.; Franke, W.W.
J. Mol. Biol. 184, 713-724, 1985
A:Title: Amino acid sequence microheterogeneities of basic (type II) cytokeratins of Xenopus laevis
A:Reference number: A2914; MUID:86011576; PMID:2413219
A:Accession: A02953
A:Molecule type: mRNA
A:Residues: 1-425 <HOF>
A:Cross-references: UNIPROT:P04266; UNIPARC:UPI000012DB06; GB:X02894; GB:M13956; NID:964
A:Experimental source: clone pUF23
C:Comment: The cytokeratin if proteins appear to be obligate heteropolymers, composed of types of proteins is low and restricted to specific isolated regions within the rod domain. Three subdomains, each characterized by a distinctive amino acid composition, d length among the intermediate filament proteins, the arrangement of its subdomains preally distant species, suggests functional significance.
C:Comment: There is only one group of closely related type II amphibian cytokeratins. The rate of differentiation of the amphibian epidermis, which lacks a true stratum corneum.
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-296/Domain: rod (fragment) <ROD>
F:1-16/Region: coil 1A (fragment)
F:17-30/Region: linker 1
F:31-131/Region: coil 1B
F:132-148/Region: linker 12
F:149-167/Region: coil 2A
F:168-175/Region: linker 2
F:176-296/Region: coil 2B
F:234/Region: stutter
F:297-425/Domain: tail <END>
F:297-405/Region: H2 and V2 subdomains
F:406-425/Region: E2 subdomain

Query Match      100.0%; Score 43; DB 1; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.4; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY      1 KYEDEINK 8
Db      72 KYEDEINK 79

RESULT 6
A43782
keratin, type II - eastern newt (fragment)
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C:Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 09-Jul-2004
C:Accession: A43782
R:Ferretti, P.; Brookes, J.P.; Brown, R.
Development 111, 497-507, 1991
A:Title: A new type II keratin restricted to normal and regenerating limbs and tails is
A:Reference number: A43782; MUID:91372147; PMID:1716554
A:Accession: A43782
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-461 <FER>
A:Cross-references: UNIPROT:O42435; UNIPARC:UPI00001774C4; GB:X57671
C:Superfamily: cytoskeletal keratin

Query Match      100.0%; Score 43; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.6; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY      1 KYEDEINK 8
Db      133 KYEDEINK 140

RESULT 7

```



A34720  
keratin 8, type II cytoskeletal - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Jul-1990 #sequence\_revision 06-Jul-1990 #text\_change 09-Jul-2004  
C:Accession: A34720; S06888; S12479; I37982; I37983; JS0487; A31942  
R:Yamamoto, R.; Kao, L.C.; McKnight, C.E.; Strauss III, J.F.  
Mol. Endocrinol. 4, 370-374, 1990  
A:Title: Cloning and sequence of cDNA for human placental cyto keratin 8. Regulation of  
A:Reference number: A34720; MUID:90258929; PMID:1692965  
A:Accession: A34720  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-483 <YAM>  
A:Cross-references: UNIPROT:P05787; UNIPARC:UPI00001774C3; GB:M34225  
R:Leube, R.E.; Bosch, F.X.; Romano, V.; Zimbelmann, R.; Hoefler, H.; Franke, W.W.  
Differentiation 33, 69-85, 1986  
A:Title: Cyto keratin expression in simple epithelia.  
A:Reference number: S06888; MUID:87134779; PMID:2434381  
A:Accession: S06888  
A:Molecule type: mRNA  
A:Residues: 205-483 <LEU>  
A:Cross-references: UNIPARC:UPI000016ABAI; EMBL:X12882  
R:Franke, W.W.  
submitted to the EMBL Data Library, September 1988  
A:Reference number: S12479  
A:Accession: S12479  
A:Molecule type: mRNA  
A:Residues: 205-309, 'I', 311-483 <FRA>  
A:Cross-references: UNIPARC:UPI000016A769; EMBL:X12882; NID:g30312; PIDN:CAA31376.1; PID  
R:Waseem, A.; Alexander, C.M.; Steel, J.B.; Lane, E.B.  
New Biol. 2, 464-478, 1990  
A:Title: Embryonic simple epithelial keratins 8 and 18: chromosomal location emphasizes  
A:Reference number: I37982; MUID:91145351; PMID:1705144  
A:Accession: I37982  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-76, 'V', 78-416, 'S', 418-428, 'G', 430-431, 'S', 433-483 <RES>  
A:Cross-references: UNIPARC:UPI0000161B3E; EMBL:X74929; NID:g400415; PIDN:CAA52882.1; PI  
A:Accession: I37983  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 151-201, 'VD', 202-271 <RES>  
A:Cross-references: UNIPARC:UPI000016ABCF; EMBL:X74981; NID:g400417; PIDN:CAA52916.1; PI  
R:Krauss, S.; Franke, W.W.  
Gene 86, 241-249, 1990  
A:Title: Organization and sequence of the human gene encoding cyto keratin 8.  
A:Reference number: JS0487; MUID:90215304; PMID:1691124  
A:Accession: JS0487  
A:Molecule type: DNA  
A:Residues: 1-76, 'V', 78-428, 'G', 'SQA', 434-483 <KRA>  
A:Cross-references: UNIPARC:UPI000016A82B; GB:M34482; NID:g181572; PIDN:AAA35763.1; PID  
R:Kulesh, D.A.; Cecena, G.; Darmon, Y.M.; Vasseur, M.; Oshima, R.G.  
Mol. Cell. Biol. 9, 1553-1565, 1989  
A:Title: Posttranslational regulation of keratins: degradation of mouse and human keratin  
A:Reference number: A31942; MUID:89261783; PMID:2471065  
A:Accession: A31942  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-76, 'V', 78-231 <KUL>  
A:Cross-references: UNIPARC:UPI000016A354; GB:M26512; NID:g177796; PIDN:AAA51542.1; PID  
C:Comment: This protein is one of the type II cyto keratins.  
C:Genetics:  
A:Gene: GDB:KRT8  
A:Cross-references: GDB:118830; OMIM:148060  
A:Map position: 12p13.2-12q24.1  
A:Introns: 108/3; 178/2; 198/3; 230/3; 327/3; 401/2; 421/1  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; intermediate filament  
Query Match 100.0%; Score 43; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
Db 178 KYEDEINK 185  
RESULT 8  
JT0407  
keratin 8, type II cytoskeletal, embryonic - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C:Accession: JT0407; A22078  
R:Ouellet, T.; Levac, P.; Royal, A.  
Gene 70, 75-84, 1988  
A:Title: Complete sequence of the mouse type-II keratin EndoA: its amino-terminal region  
A:Reference number: JT0407; MUID:89196919; PMID:2467842  
A:Accession: JT0407  
A:Molecule type: mRNA  
A:Residues: 1-487 <OUE>  
A:Cross-references: UNIPROT:Q61518; UNIPARC:UPI0000086EEO; GB:M22831; NID:G511653; PIDN:  
A:Experimental source: carcinoma cell line F9-21  
R:Vasseur, M.; Duprey, P.; Brulet, P.; Jacob, P.  
Proc. Natl. Acad. Sci. U.S.A. 82, 1155-1159, 1985  
A:Title: One gene and one pseudogene for the cyto keratin endo A.  
A:Reference number: A22078; MUID:85140247; PMID:2579387  
A:Accession: A22078  
A:Molecule type: DNA  
A:Residues: 1-13 <VAS>  
A:Cross-references: UNIPARC:UPI000016CD31; GB:K02783; NID:g193019; PIDN:AAA37549.1; PID:  
C:Genetics:  
A:Gene: EndoA  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; intermediate filament  
F:1-94/Domain: head <HED>  
F:95-406/Domain: rod <ROD>  
F:129-146/Region: linker 1  
F:147-235/Region: coil 1B  
F:236-260/Region: linker 12  
F:261-406/Region: coil 2  
F:407-487/Domain: tail <TAI>  
Query Match 100.0%; Score 43; DB 2; Length 487;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
Db 181 KYEDEINK 188  
RESULT 9  
JS0658  
cyto keratin EndoA - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C:Accession: JS0658  
R:Tamai, Y.; Takemoto, Y.; Matsumoto, M.; Morita, T.; Matsushiro, A.; Nozaki, M.  
Gene 104, 169-176, 1991  
A:Title: Sequence of the EndoA gene encoding mouse cyto keratin and its methylation state  
A:Reference number: JS0658; MUID:92009210; PMID:1717348  
A:Accession: JS0658  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-490 <TAM>  
A:Cross-references: UNIPROT:Q61463; UNIPARC:UPI00000220F8; GB:D90360; NID:g220391; PIDN:  
C:Comment: This type-II cyto keratin forms intermediate filaments together with type-I cy  
C:Genetics:  
A:Gene: EndoA  
A:Introns: 114/3; 184/2; 204/3; 236/3; 291/3; 333/3; 407/2; 427/1  
C:Superfamily: cytoskeletal keratin  
Query Match 100.0%; Score 43; DB 2; Length 490;  
Best Local Similarity 100.0%; Pred. No. 1.7;

```
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
    |||||||
Db 184 KYEDEINK 191

RESULT 10
S08381
keratin, 58K type II, cytoskeletal (clone pXenCK55(5/6)) - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 09-Jul-2004
C:Accession: S08381
R:Fouquet, B.; Herrmann, H.; Franz, J.K.; Franke, W.W.
Development 104, 533-548, 1988
A:Title: Expression of intermediate filament proteins during development of Xenopus laevis
A:Reference number: S08381; MUID:90032362; PMID:2478354
A:Accession: S08381
A:Molecule type: mRNA
A:Residues: 1-513 <FOU>
A:Cross-references: UNIPROT:P16878; UNIPARC:UPI000017151D; EMBL:X14427; NID:964625; PIDN:
C:Superfamily: cytoskeletal keratin

Query Match 100.0%; Score 43; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
    |||||||
Db 216 KYEDEINK 223

RESULT 11
JS0291
intermediate filament protein ON3 - goldfish
C:Species: Carassius auratus (goldfish)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jul-2004
C:Accession: JS0291
R:Giordano, S.; Glasgow, E.; Tesser, P.; Schechter, N.
Neuron 2, 1507-1516, 1989
A:Title: A type II keratin is expressed in glial cells of the goldfish visual pathway.
A:Reference number: JS0291; MUID:90180470; PMID:2483326
A:Accession: JS0291
A:Molecule type: mRNA
A:Residues: 1-520 <GIO>
A:Cross-references: UNIPROT:P18520; UNIPARC:UPI000012D74A
A:Experimental source: Glial cell
C:Comment: This protein is one of the nonneuronal predominant intermediate filament prot
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; optic nerve; visual pathway
F:1-108/Domain: head <HEA>
F:109-144/Region: coil 1A
F:145-157/Region: linker
F:158-253/Region: coil 1B
F:254-273/Region: linker
F:274-420/Region: coil 2
F:421-520/Domain: tail <TAI>

Query Match 100.0%; Score 43; DB 2; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
    |||||||
Db 196 KYEDEINK 203

RESULT 12
I59009
epidermal keratin subunit II - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I59009; A61205
```

```
R:Steinert, P.M.; Parry, D.A.D.; Racoosin, E.L.; Idler, W.W.; Steven, A.C.; True, B.L.; F
Proc. Natl. Acad. Sci. U.S.A. 81, 5709-5713, 1984
A:Title: The complete cDNA and deduced amino acid sequence of a type II mouse epidermal
A:Reference number: I59009; MUID:85014838; PMID:6207530
A:Accession: I59009
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-553 <RES>
A:Cross-references: UNIPROT:P50446; UNIPARC:UPI000016C887; GB:K02108; NID:9198634; PIDN:
R:Finch, J.; Andrews, K.; Krieg, P.; Fuerstenberger, G.; Slaga, T.; Ootsuyama, A.; Tanoo
Carcinogenesis 12, 1519-1522, 1991
A:Title: Identification of a cloned sequence activated during multi-stage carcinogenesis
A:Reference number: A61205; MUID:91316763; PMID:1713533
A:Accession: A61205
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 528-553 <PIN>
A:Cross-references: UNIPARC:UPI00001774C8
C:Genetics:
A:Gene: KER2
C:Superfamily: cytoskeletal keratin

Query Match 100.0%; Score 43; DB 2; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
    |||||||
Db 241 KYEDEINK 248

RESULT 13
KXUEA
keratin 6a, type II - human
N:Alternate names: 56-kDa type II keratin; keratin cytoskeletal
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: A57398; A02944
R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.
J. Biol. Chem. 270, 18581-18592, 1995
A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly
A:Reference number: A57398; MUID:95355491; PMID:7543104
A:Accession: A57398
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-564 <TKX>
A:Cross-references: UNIPROT:P02538; UNIPARC:UPI000013CD4C; GB:L42575; NID:9908769; GB:L4;
74; GB:L42581; NID:9908775; GB:L42582; NID:9908776; GB:L42583; NID:9908777; PIDN:AAC4176;
R:Hanukoglu, I.; Fuchs, E.
Cell 33, 915-924, 1983
A:Title: The cDNA sequence of a type II cytoskeletal keratin reveals constant and variabl
A:Reference number: A02944; MUID:83259278; PMID:6191871
A:Accession: A02944
A:Molecule type: mRNA
A:Residues: 208-394, 'S', 396-564 <HAN>
A:Cross-references: UNIPARC:UPI000016AB87; GB:J00269; NID:934068; PIDN:CAA24760.1; PID:gi:
C:Comment: The cytoskeletal and microfilillar keratins are classified into two types, t
atin IF protein subunit appears to be a heterotetramer of two type I and two type II prot
C:Genetics:
A:Gene: GDB:KRT6A
A:Cross-references: GDB:128111; OMIM:148041
A:Map position: 12q12-12q21
A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:2-163/Domain: head <HED>
F:2-126/Region: E1 and V1 subdomains
F:127-163/Region: H1 subdomain
F:162-476/Domain: rod <ROD>
F:199-210/Region: linker 1
F:211-311/Region: coil 1B
F:312-328/Region: linker 12
F:329-347/Region: coil 2A
```

F:348-355/Region: linker 2  
F:356-476/Region: coil 2B  
F:414/Region: stutter  
F:477-564/Domain: tail <END>  
F:477-496/Region: H2 subdomain  
F:497-564/Region: V2 and E2 subdomains

Query Match 100.0%; Score 43; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||||

Db 252 KYEDEINK 259

## RESULT 14

KRUHUB

keratin 6b, type II - human

N:Alternate names: 56K type II keratin; keratin, cytoskeletal

C:Species: Homo sapiens (man)

C&gt;Date: 04-Dec-1986 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: I61767; A02945

R:Takahashi, K.; Paladini, R.D.; Coulombe, P.A.

J. Biol. Chem. 270, 18581-18592, 1995

A:Title: Cloning and characterization of multiple human genes and cDNAs encoding highly

A:Reference number: A57398; MUID:95355491; PMID:7543104

A:Accession: I61767

A&gt;Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-564 &lt;RES&gt;

A:Cross-references: UNIPROT:P04259; UNIPARC:UPI0000167B38; GB:L42594; NID:g908780; GB:L42594; GB:L42590; NID:g908786; GB:L42591; NID:g908787; GB:L42592; NID:g908788; PIDN:AAAC4176

R:Tyner, A.L.; Eichman, M.J.; Fuchs, E.

Proc. Natl. Acad. Sci. U.S.A. 82, 4683-4687, 1985

A:Title: The sequence of a type II keratin gene expressed in human skin: conservation of

A:Reference number: A02945; MUID:85270392; PMID:2410904

A:Accession: A02945

A:Molecule type: DNA

A:Residues: 2-88, 'AG', '91-115, 'PA', '118, 'LC', '122-158, 'IG', '161-254, 'V', '256-564 &lt;TYN&gt;

A:Cross-references: UNIPARC:UPI0000173D58; GB:M11229; GB:L00205; NID:g186714; PIDN:AAA59

A&gt;Note: Initiator Met not shown

C:Comment: There are two types of cytoskeletal and microfibrillar keratin, I (acidic) and

biligate heteropolymers, composed of complexes formed by the aggregation of at least one

C:Genetics:

A:Gene: GDB:KRT6B

A:Cross-references: GDB:128113; OMIM:148042

A:Map position: 12pter-12qter

A:Introns: 180/3; 252/2; 272/3; 304/3; 359/3; 401/3; 475/2; 487/1

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

F:2-163/Domain: head &lt;HED&gt;

F:2-127/Region: E1 and V1 subdomains

F:128-163/Region: H1 subdomain

F:164-476/Domain: rod &lt;ROD&gt;

F:164-198/Region: coil 1A

F:199-210/Region: linker 1

F:211-311/Region: coil 1B

F:312-328/Region: linker 12

F:329-347/Region: coil 2A

F:348-355/Region: linker 2

F:356-476/Region: coil 2B

F:414/Region: stutter

F:477-564/Domain: tail &lt;END&gt;

F:477-496/Region: H2 subdomain

F:497-564/Region: V2 and E2 subdomains

Query Match 100.0%; Score 43; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||||

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 81.8641 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-12  
Perfect score: 43  
Sequence: 1 KYEDEINK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	176	Q4J123 ICTPU	Q4j123 ictalurus p
2	43	100.0	222	Q4L0E7 MOUSE	Q4l0e7 mus musculus
3	43	100.0	224	Q678L1 MOUSE	Q678l1 mus musculus
4	43	100.0	310	Q28810 POTTR	Q28810 potorinus tr
5	43	100.0	324	Q80VP7 MOUSE	Q80vp7 mus musculus
6	43	100.0	346	Q4QV72 SPAAU	Q4qv72 sparus aura
7	43	100.0	365	Q96910 HUMAN	Q96910 homo sapien
8	43	100.0	384	K2C6D HUMAN	P48667 homo sapien
9	43	100.0	417	Q4SRP0 TETNG	Q4srp0 tetraodon n
10	43	100.0	419	K2C2 XENLA	P04265 xenopus lae
11	43	100.0	425	K2C1 XENLA	P04266 xenopus lae
12	43	100.0	461	Q42435 NOTVI	O42435 notophthalm
13	43	100.0	482	K2C8 HUMAN	P05787 homo sapien
14	43	100.0	482	K2C8 RAT	Q10758 rattus norv
15	43	100.0	483	Q53GJ0 HUMAN	Q53gj0 homo sapien
16	43	100.0	483	Q6DHW5 HUMAN	Q6dhw5 homo sapien
17	43	100.0	483	Q6GMY0 HUMAN	Q6gmy0 homo sapien
18	43	100.0	483	Q5RB34 PONY	Q5rb34 pongo pygma
19	43	100.0	483	Q80WH8 MURI	Q80wh8 rattus sp.
20	43	100.0	486	Q9YH07 9CHON	Q9yh07 scyllorhinu
21	43	100.0	487	Q6L518 MOUSE	Q6l518 mus musculus
22	43	100.0	487	Q7SYF6 ACIBE	Q7syf6 acipenser b
23	43	100.0	490	Q3KOK5 MOUSE	Q3kdk5 mus musculus
24	43	100.0	490	Q3TGI1 MOUSE	Q3tgi1 mus musculus
25	43	100.0	490	Q3TJEL MOUSE	Q3tjel mus musculus
26	43	100.0	490	Q3TKI7 MOUSE	Q3tki7 m blascocys
27	43	100.0	490	Q61463 MOUSE	Q61463 gallus gall
28	43	100.0	492	K2C0 CHICK	Q42434 notophthalm
29	43	100.0	496	Q42434 NOTVI	Q42434 notophthalm
30	43	100.0	497	Q5K2N3 PROAT	Q5k2n3 prototenus
31	43	100.0	497	Q6N160 BRARE	Q6n160 brachydanio

32	43	100.0	498	2	Q9PV92 BRARE	Q9pv92 brachydanio
33	43	100.0	499	2	Q9H552 HUMAN	Q9h552 homo sapien
34	43	100.0	499	2	Q7ZT78 BRARE	Q7zt78 brachydanio
35	43	100.0	505	2	Q6P4C7 HUMAN	Q6p4c7 homo sapien
36	43	100.0	508	2	Q6NVR6 XENTR	Q6nvrf xenopus tro
37	43	100.0	511	2	Q90ZF7 RANCA	Q90zf7 rana cateab
38	43	100.0	512	1	K2C5 XENLA	P16878 xenopus lae
39	43	100.0	515	2	Q8JFG4 ONCMY	Q8jfg4 oncorhynch
40	43	100.0	519	1	K2C1B RAT	Q61901 rattus norv
41	43	100.0	520	1	ION3 CARAU	P18520 carassius a
42	43	100.0	520	2	Q6NWF6 BRARE	Q6nwf6 brachydanio
43	43	100.0	521	2	Q5K2N4 PROAT	Q5k2n4 prototenus
44	43	100.0	521	2	Q5K2N5 PROAT	Q5k2n5 prototenus
45	43	100.0	529	2	Q919P5 RANCA	Q919p5 rana cateab

ALIGNMENTS

RESULT 1  
Q4J123 ICTPU PRELIMINARY; PRT; 176 AA.  
AC Q4J123;  
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 02-AUG-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Keratin 8 (Fragment).  
OS Ictalurus punctatus (Channel catfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Siluriformes;  
OC Ictaluridae; Ictalurus.  
OX NCBI\_TaxID=7998;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Yeh H.-Y., Klesius P.H.;  
RT "Differential gene expression of Ictalurus punctatus after  
Edwardsiella ictaluri infection";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; DQ086174; AAY86957.1; -; mRNA.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002957; Keratin I.  
DR InterPro; IPR003054; Keratin II.  
DR PANTHER; PTHR1893.SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PRINTS; PR01276; TYPE2KERATIN.  
KW Keratin.  
FT NON\_TER 176  
FT NON\_TER 176  
SQ SEQUENCE 176 AA; 20651 MW; 8EE3A88097529C33 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
Db 69 KYEDEINK 76

RESULT 2  
Q4L0E7 MOUSE PRELIMINARY; PRT; 222 AA.  
ID Q4L0E7; MOUSE  
AC Q4L0E7;  
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 02-AUG-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Type II cyokeratin Kb39 (Fragment).

```
GN Name=4732484G22Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NMRI; TISSUE=Tail skin;
RA Zimek A., Hesse M.;
RT "Cornea-specific cytochrome K3 seems to be absent in the rodent
lineage.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY651826; AAV85505.1; -; mRNA.
DR MGI; MGI:3588209; 4732484G22Rik.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002957; Keratin_1.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
FT NON_TER 1 222
FT NON_TER 222
SQ SEQUENCE 222 AA; 25801 MW; 483B4F589032C1D1 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 222;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDEINK 8
Db |||||||
69 KYDEINK 76

RESULT 3
Q578L1_MOUSE
ID Q678L1_MOUSE PRELIMINARY; PRT; 224 AA.
AC Q678L1.
DT 11-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Keratin 1b (Fragment).
GN Name=4732484G22Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NMRI; TISSUE=Tail skin;
RA Zimek A., Hesse M., Magin T., Weber K.;
RT "Homologies between the human, mouse and rat keratin-clusters.";
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY379112; AAR32930.1; -; mRNA.
DR Ensembl; ENSMUSG0000067594; Mus musculus.
DR MGI; MGI:3588209; 4732484G22Rik.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.

Query Match 100.0%; Score 43; DB 2; Length 310;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDEINK 8
Db |||||||
179 KYDEINK 186

RESULT 5
Q80VP7_MOUSE
ID Q80VP7_MOUSE PRELIMINARY; PRT; 324 AA.
AC Q80VP7.
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
PANTHER; PTHR1893:SF5; Keratin_II; 1.
Pfam; PF00038; Filament; 1.
PRINTS; PR01248; TYPE1KERATIN.
PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
FT NON_TER 1 310
FT NON_TER 310
SQ SEQUENCE 310 AA; 35170 MW; 31D37E0BC63BFA85 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 310;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDEINK 8
Db |||||||
179 KYDEINK 186

RESULT 4
Q28810_POTTR
ID Q28810_POTTR PRELIMINARY; PRT; 310 AA.
AC Q28810.
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Keratin 8 (Fragment).
OS Porcous tridactylus (Potoroo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Potoroidae; Potorous.
OX NCBI_TaxID=9310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94118290; PubMed=7507169;
RA Boettger V., Lane E.B.;
RT "A monoclonal antibody Epitope on Keratin 8 identified using a phage
peptide library.";
RL J. Mol. Biol. 235:61-67(1994).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; X70987; CAA50316.1; -; mRNA.
DR PIR; S43865; S43865.
DR HSSP; P08670; 1GK7.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR InterPro; IPR003054; Keratin_II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
KW Keratin.
FT NON_TER 1 310
FT NON_TER 310
SQ SEQUENCE 310 AA; 35170 MW; 31D37E0BC63BFA85 CRC64;
```

DT 07-FEB-2006, entry version 20.  
DE Hypothetical protein MGC54654.  
GN Name=MGC54654;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Eye;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., McKernan K.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., Peters K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Eye;  
RG NIH MGC Project;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC  
DR EMBL; BC046626; AAH46626.1; -; mRNA.  
DR HSSP; P08670; 1GK7.  
DR Ensembl; ENSMUSG00000046162; Mus musculus.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR003054; Keratin II.  
DR PANTHER; PTHR18893.SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 324 AA; 35165 MW; 0E4D7E25144435DD CRC64;  
Query Match 100.0%; Score 43; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KYEDEINK 8  
DB 202 KYEDEINK 209  
-----  
RESULT 6  
Q4QY72 SPAAU PRELIMINARY; PRT; 346 AA.  
ID Q4QY72  
AC Q4QY72  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 6.  
DE Type II keratin E3-like protein.  
OS Sparus aurata (Gilthead sea bream).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
OC Sparidae; Sparus.  
OX NCBI\_TaxID=8175;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Sarrapoulou E., Power D.M., Magoulas A., Geisler R., Koroulas G.;  
RT "Comparative analysis and characterization of expressed sequence tags  
in gilthead sea bream (Sparus aurata) liver and embryos.";  
RL Aquaculture 243:69-81 (2005).  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC  
DR EMBL; AY550955; AAT44423.1; -; mRNA.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR InterPro; IPR003054; Keratin\_II.  
DR PANTHER; PTHR18893.SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 346 AA; 38624 MW; 75468AFEC8048F5 CRC64;  
Query Match 100.0%; Score 43; DB 2; Length 346;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KYEDEINK 8  
DB 61 KYEDEINK 68  
-----  
RESULT 7  
Q96910 HUMAN PRELIMINARY; PRT; 365 AA.  
ID Q96910  
AC Q96910;  
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
DT 24-MAY-2005, sequence version 2.  
DT 07-FEB-2006, entry version 23.  
DE KRT8 protein (fragment).  
GN Name=KRT8;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters K.J., Malek J.A., Gunaratne P.H.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullahy S.J.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human

```

RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
RN [2]
RC NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RG NIH MGC Project;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BC011373; AAH11373.2; -; mRNA.
DR Ensembl; ENSG00000170421; Homo sapiens.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin I.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament.
FT NON TER 1
FT SEQUENCE 365 AA; 41108 MW; 57CD83091D0635E8 CRC64;
SQ
Query Match 100.0%; Score 43; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KYDEINK 8
DB 60 KYDEINK 67
RESULT 8
K2C6D HUMAN
ID K2C6D HUMAN STANDARD; PRT; 384 AA.
AC P48667;
DT 01-FEB-1996, integrated into UniProtKB/Swiss-Prot.
DT 01-FEB-1996, sequence version 1.
DT 07-FEB-2006, entry version 44.
DE Keratin, type II cytoskeletal 6D (Cytokeratin-6D) (CK 6D) (K6d
DE keratin) (Fragment).
GN Name=KRT6D;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC TISSUE=Skin;
RX MEDLINE=9535491; PubMed=7543104; DOI=10.1074/jbc.270.31.18581;
RA Takahashi K., Paladini R.D., Coulombe P.A.;
RT "Cloning and characterization of multiple human genes and cDNAs
RT encoding highly related type II keratin 6 isoforms.";
RL J. Biol. Chem. 270:18581-18592(1995).
CC -!- FUNCTION: There are two types of cytoskeletal and microfibrillar
CC keratin: I (acidic; 40-55 kDa) [K9 to K20] and II (neutral to
CC basic; 56-70 kDa) [K1 to K8]. Both a basic and an acidic keratin
CC are required for filament assembly.
CC -!- SUBUNIT: Heterodimer of a type I and a type II keratin. KRT6
CC isomers associate with KRT16 and/or KRT17.
CC -!- TISSUE SPECIFICITY: Constitutively expressed in distinct types of
CC epithelia such as those in oral mucosa, esophagus, papillae of
CC tongue and hair follicle outer root sheath.
CC -!- MISCELLANEOUS: There are at least six isoforms of human type II
CC keratin 6 (K6).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----

```

```

CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; L42610; AAB60696.1; -; Genomic DNA.
DR EMBL; L42603; AAB60696.1; JOINED; Genomic DNA.
DR EMBL; L42604; AAB60696.1; JOINED; Genomic DNA.
DR EMBL; L42605; AAB60696.1; JOINED; Genomic DNA.
DR EMBL; L42606; AAB60696.1; JOINED; Genomic DNA.
DR EMBL; L42607; AAB60696.1; JOINED; Genomic DNA.
DR EMBL; L42608; AAB60696.1; JOINED; Genomic DNA.
DR EMBL; L42609; AAB60696.1; JOINED; Genomic DNA.
DR PIR; I61769; I61769.
DR HSSP; P08670; IGK4.
DR Ensembl; ENSG00000170455; Homo sapiens.
DR HGNC; HGNC:20405; KRT6D.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; NAS.
DR GO; GO:0030154; P:cell differentiation; NAS.
DR GO; GO:0008284; P:positive regulation of cell proliferation; NAS.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR1893:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament; Keratin.
FT CHAIN <1 384
FT SEQUENCE 384 AA; 42468 MW; D69DCFD808C94E91 CRC64;
SQ
Query Match 100.0%; Score 43; DB 1; Length 384;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KYDEINK 8
DB 72 KYDEINK 79
RESULT 9
Q4SRP0 TETNG
ID Q4SRP0 TETNG PRELIMINARY; PRT; 417 AA.
AC Q4SRP0;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Chromosome undetermined SCAF14505, whole genome shotgun sequence.
OS ORFNames=GSTENG0013817001;
GN Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=1549614; DOI=10.1038/nature03025;
RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Farra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,

```



RA Kellis M., Volf J.-N., Guigo R., Zody M.C., Mesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Winkler P., Lander E.S., Weissenbach J., Roest Crolius H.;  
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 RL the early vertebrate proto-karyotype.";  
 RN Nature 431:946-957(2004).  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; CAAE01014505; CAF96692.1; -; Genomic DNA.  
 DR GO; GO:0005882; C:intermediate filament; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR18893.SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 DR Intermediate filament.  
 KW SEQUENCE 419 AA; 46557 MW; B2F4D4E29E625589 CRC64;  
 SQ

Query Match 100.0%; Score 43; DB 2; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDEINK 8  
 |||||  
 Db 221 KYDEINK 228

RESULT 10  
 K2C2 XENLA  
 ID K2C2 XENLA STANDARD; PRT; 419 AA.  
 AC P04255;  
 DT 20-MAR-1987, integrated into UniProtKB/Swiss-Prot.  
 DT 07-FEB-2006, entry version 39.  
 DE Keratin, type II cytoskeletal I (Clone PUF164) (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RX MEDLINE=86011576; PubMed=2413219;  
 RA Hoffmann W., Franz J.K., Franke W.W.;  
 RT "Amino acid sequence microheterogeneities of basic (type II)  
 RT cytokeratins of Xenopus laevis epidermis and evolutionary  
 RT conservativity of helical and non-helical domains.";  
 RL J. Mol. Biol. 184:713-724(1985).  
 CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
 CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
 CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
 CC 55 and 56-70 kilodaltons, respectively).  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; X02895; CAA26654.1; -; mRNA.  
 DR PIR; A02952; KRXL2B.  
 DR HSSP; P08670; 1GK4.

DR InterPro; IPR001664; IF.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR18893.SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Coiled coil; Intermediate filament; Keratin.  
 FT CHAIN <1 419  
 FT Keratin, type II cytoskeletal I.  
 FT /FTId=PRO\_0000063719.  
 FT REGION <1 271 Rod.  
 FT REGION <1 16 Linker 1.  
 FT REGION 17 108 Coiled 1B.  
 FT REGION 109 132 Linker 12.  
 FT REGION 133 271 Coiled 2.  
 FT REGION 272 419 Tail.  
 FT SITE 215 215 Stutter.  
 FT NON TER 1 1  
 SQ SEQUENCE 419 AA; 44595 MW; 15D79584FC31DF15 CRC64;

Query Match 100.0%; Score 43; DB 1; Length 419;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYDEINK 8  
 |||||  
 Db 51 KYDEINK 58

RESULT 11  
 K2C1 XENLA  
 ID K2C1 XENLA STANDARD; PRT; 425 AA.  
 AC P04266;  
 DT 20-MAR-1987, integrated into UniProtKB/Swiss-Prot.  
 DT 20-MAR-1987, sequence version 1.  
 DT 07-FEB-2006, entry version 39.  
 DE Keratin, type II cytoskeletal I (Clone PUF23) (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RX MEDLINE=86011576; PubMed=2413219;  
 RA Hoffmann W., Franz J.K., Franke W.W.;  
 RT "Amino acid sequence microheterogeneities of basic (type II)  
 RT cytokeratins of Xenopus laevis epidermis and evolutionary  
 RT conservativity of helical and non-helical domains.";  
 RL J. Mol. Biol. 184:713-724(1985).  
 CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
 CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
 CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
 CC 55 and 56-70 kilodaltons, respectively).  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; X02894; CAA26653.1; -; mRNA.  
 DR PIR; A02953; KRXL2A.  
 DR HSSP; P08670; 1GK4.

DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin I.  
 DR InterPro; IPR003054; Keratin II.  
 DR PANTHER; PTHR18893.SF5; Keratin\_II; 1.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Coiled coil; Intermediate filament; Keratin.  
 FT CHAIN <1 425  
 FT Keratin, type II cytoskeletal I.  
 FT /FTId=PRO\_0000063718.  
 FT REGION <1 292 Rod.

```
FT REGION <1 16 Coil 1A.
FT REGION 17 37 Linker 1.
FT REGION 38 129 Coil 1B.
FT REGION 130 153 Linker 12.
FT REGION 154 292 Coil 2.
FT REGION 293 425 Tail.
FT SITE 236 236 Stutter.
FT NON TER 1
SQ SEQUENCE 425 AA; 45725 MW; D58D15DEC8C1C4E3 CRC64;

Query Match 100.0%; Score 43; DB 1; Length 425;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
DB 72 KYEDEINK 79

RESULT 12
ID O42435 NOTVI PRELIMINARY; PRT; 461 AA.
AC O42435;
DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.
DT 01-JAN-1998, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Cytokeratin type II (Fragment).
GN Name=KII;
OS Notoththalmus viridescens (Eastern newt) (Triturus viridescens).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae;
OC Notoththalmus.
OX NCBI_TaxID=8316;
RN [1]_TaxID=8316;
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Limb blastema;
RX MEDLINE=91372147; PubMed=1716554;
RA Ferretti P., Brookes J.P., Brown R.;
RT "A new type II keratin restricted to normal and regenerating limbs
and tails is responsive to retinoic acid.";
RL Development 111:497-507(1991).
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Limb blastema;
RA Ferretti P.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ001295; CAA04655.1; -; mRNA.
DR PIR; A43782; A43782.
DR HSP; P08670; IGK7.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin II.
DR PANTHER; PTHR18993:SF5; Keratin_II; 1.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
FT NON TER 1
SQ SEQUENCE 461 AA; 50638 MW; F3A018A7C10CA0FD CRC64;

Query Match 100.0%; Score 43; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
DB 133 KYEDEINK 140
```

```
RESULT 13
K2C8 HUMAN
ID K2C8 HUMAN STANDARD; PRT; 482 AA.
AC P05787; Q14099; Q14716; Q14717; Q96J60;
DT 01-NOV-1988, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 5.
DT 07-FEB-2006, entry version 72.
DE Keratin, type II cytoskeletal 8 (Cytokeratin-8) (CK-8) (Keraton-8) (K8).
GN Name=KRT8; Synonym=CYK8;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=90258929; PubMed=1692965;
RY Yamamoto R., Kao L.C., McKnight C.E., Strauss J.F. III;
RT "Cloning and sequence of cDNA for human placental cyokeratin 8.
Regulation of the mRNA in trophoblastic cells by cAMP.";
RL Mol. Endocrinol. 4:370-374(1990).
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=90215304; PubMed=1691124; DOI=10.1016/0378-1119(90)90285-Y;
RX Krauss S., Franke W.W.;
RT "Organization and sequence of the human gene encoding cyokeratin 8.";
RL Gene 86:241-249(1990).
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=91145351; PubMed=1705144;
RX Waseem A., Alexander C.M., Steel J.B., Lane E.B.;
RT "Embryonic simple epithelial keratins 8 and 18: chromosomal location
emphasizes difference from other keratin pairs.";
RL New Biol. 2:464-478(1990).
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=97207326; PubMed=9054461; DOI=10.1074/jbc.272.11.7556;
RX Ku N.-O., Omary M.B.;
RT "Phosphorylation of human keratin 8 in vivo at conserved head domain
serine 23 and at epidermal growth factor-stimulated tail domain serine
431.";
RL J. Biol. Chem. 272:7556-7564(1997).
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RY Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Colling P.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN NUCLEOTIDE SEQUENCE OF 1-230.
RX MEDLINE=89261783; PubMed=2471065;
RY Kulesh D.A., Cecena G., Darmon Y.M., Vasseur M., Oshima R.G.;
```

RT "Posttranslational regulation of keratins: degradation of mouse and  
human keratins 18 and 8.";  
RL Mol. Cell. Biol. 9:1553-1565(1989).  
RN [7]  
RX NUCLEOTIDE SEQUENCE OF 204-482.  
RP MEDLINE=87134779; PubMed=2434381;  
RA Leube R.E., Bosch F.X., Romano V., Zimbelmann R., Hofer H.,  
RA Franke W.W.;  
RT "Cyokeratin expression in simple epithelia. III. Detection of mRNAs  
encoding human cyokeratins nos. 8 and 18 in normal and tumor cells by  
hybridization with cDNA sequences in vitro and in situ.";  
RL Differentiation 33:69-85(1986).  
RN [8]  
RN PARTIAL PROTEIN SEQUENCE.  
RP TISSUE=Colon carcinoma;  
RX MEDLINE=97295306; PubMed=9150948;  
RA Ji H., Reid G.E., Moritz R.L., Eddes J.S., Burgess A.W., Simpson R.J.;  
RT "A two-dimensional gel database of human colon carcinoma proteins.";  
RL Electrophoresis 18:605-613(1997).  
RN [9]  
RN INTERACTION WITH PNN.  
RP PubMed=10809736; DOI=10.1074/jbc.275.20.14910;  
RA Shi J., Sugrue S.P.;  
RT "Dissection of protein linkage between keratins and pinin, a protein  
with dual location at desmosome-intermediate filament complex and in  
the nucleus.";  
RL J. Biol. Chem. 275:14910-14915(2000).  
RN [10]  
RN PHOSPHORYLATION SITE SER-73.  
RX MEDLINE=21909496; PubMed=11781324; DOI=10.1074/jbc.M111436200;  
RA He T., Stepulak A., Holmstrom T.H., Omary M.B., Eriksson J.E.;  
RT "The intermediate filament protein keratin 8 is a novel cytoplasmic  
substrate for c-Jun N-terminal kinase.";  
RL J. Biol. Chem. 277:10767-10774(2002).  
RN [11]  
RN PHOSPHORYLATION SITE SER-73, AND MUTAGENESIS OF LEU-71 AND SER-73.  
RX MEDLINE=21909510; PubMed=11788593; DOI=10.1074/jbc.M107623200;  
RA Ku N.O., Azhar S., Omary M.B.;  
RT "Keratin 8 phosphorylation by p38 kinase regulates cellular keratin  
filament reorganization: modulation by a keratin 1-like disease  
causing mutation.";  
RL J. Biol. Chem. 277:10775-10782(2002).  
RN [12]  
RN INTERACTION WITH HCV CORE PROTEIN.  
RX PubMed=15846844; DOI=10.1002/pmic.200401093;  
RA Kang S.-M., Shin M.-J., Kim J.-H., Oh J.-W.;  
RT "Proteomic profiling of cellular proteins interacting with the  
hepatitis C virus core protein.";  
RL Proteomics 5:2227-2237(2005).  
RN [13]  
RN VARIANTS CRYPTOGENIC CIRRHOSIS VAL-52; CYS-53 AND CYS-61, AND VARIANT  
VAL-62.  
RX PubMed=12724528; DOI=10.1073/pnas.09361651100;  
RA Ku N.-O., Darling J.M., Krams S.M., Esquivel C.O., Keefe E.B.,  
RA Sibley R.K., Lee Y.M., Wright T.L., Omary M.B.;  
RT "Keratin 8 and 18 mutations are risk factors for developing liver  
disease of multiple etiologies.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:6063-6068(2003).  
CC -1- SUBUNIT: Heterotetramer of two type I and two type II keratins.  
CC Keratin 8 associates with keratin 18. Interacts with HCV core  
CC protein and PNN.  
CC -1- PTM: Phosphorylation on serine residues is enhanced during EGF  
CC stimulation and mitosis. Ser-73 phosphorylation plays an important  
CC role in keratin filament reorganization.  
CC -1- DISEASE: Defects in KRT8 are a cause of cryptogenic cirrhosis  
CC [MIM:215600].  
CC -1- MISCELLANEOUS: There are two types of cytoskeletal and  
CC microfibrillar keratin: I (acidic; 40-55 kDa) [K9 to K20] and II  
CC (neutral to basic; 56-70 kDa) [K1 to K8].  
CC -1- SIMILARITY: Belongs to the intermediate filament family.  
CC -1- DATABASE: NAME=Human Intermediate Filament Mutation Database;  
CC WWW="http://www.interfil.org".  
CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: M34225; AAA35748.1; -; mRNA.  
CC EMBL: M34482; AAA35763.1; -; Genomic\_DNA.  
CC EMBL: X74929; CAAS2882.1; -; mRNA.  
CC EMBL: U76549; AAB18966.1; -; mRNA.  
CC EMBL: X98614; CAAS67203.1; -; mRNA.  
CC EMBL: M26512; AAA51542.1; -; mRNA.  
CC EMBL: BC000654; AAH00654.2; ALT INIT; mRNA.  
CC EMBL: X12882; CAA31376.1; -; mRNA.  
CC EMBL: X74981; CAAS2916.1; -; Genomic\_DNA.  
CC PIR: A34720; A34720.  
CC HSP; P08670; IGK7.  
CC IntAct; P05787; -.  
CC GlycoSuiteDB; P05787; -.  
CC SWISS-2DPAGE; P05787; HUMAN.  
CC Siens-2DPAGE; P05787; -.  
CC Ensembl; ENSG00000170421; Homo sapiens.  
CC HGNC; HGNC:6446; KRT8.  
CC MIM; 148060; gene.  
CC LinkHub; P05787; -.  
CC GO; GO:0005882; C:intermediate filament; NAS.  
CC GO; GO:0005196; F:structural molecule activity; NAS.  
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; NAS.  
CC InterPro; IPR001664; IF.  
CC InterPro; IPR002957; Keratin I.  
CC InterPro; PTHR003054; Keratin II.  
CC PANTHER; PTHR1893.SF5; Keratin\_II, 1.  
CC Pfam; PF00038; Filament; 1.  
CC PRINTS; PR01248; TYPE1KERATIN.  
CC PRINTS; PR01276; TYPE2KERATIN.  
CC PROSITE; PS00226; IF, 1.  
CC Coiled coil; Direct protein sequencing; Disease mutation;  
KW Intermediate filament; Keratin; Phosphorylation; Polymorphism.  
FT INIT\_MET 0 0  
FT CHAIN 1 482 Keratin, type II cytoskeletal 8.  
FT REGION 1 89 Head.  
FT REGION 90 397 Rod.  
FT REGION 90 125 Coil 1A.  
FT REGION 126 142 Linker 1.  
FT REGION 143 234 Coil 1B.  
FT REGION 235 258 Linker 12.  
FT REGION 259 397 Coil 2.  
FT REGION 260 381 Necessary for interaction with PNN.  
FT REGION 398 482 Tail.  
FT SITE 341 341 Stutter.  
FT MOD\_RES 23 23 Phosphoserine.  
FT MOD\_RES 73 73 Phosphoserine (by MAPK).  
FT MOD\_RES 431 431 Phosphoserine (by CAMK2 and MAPK).  
FT VARIANT 52 52 G -> V (in cryptogenic cirrhosis).  
FT VARIANT 53 53 /FTId=VAR\_023058.  
FT VARIANT 53 53 Y -> C (in cryptogenic cirrhosis).  
FT VARIANT 61 61 /FTId=VAR\_023059.  
FT VARIANT 61 61 G -> C (in cryptogenic cirrhosis).  
FT VARIANT 61 61 /FTId=VAR\_023060.  
Query Match 100.0%; Score 43; DB 1; Length 482;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KYEIDEINK 8  
Db 177 KYEIDEINK 184  
RESULT 14  
K2C8\_RAT  
ID K2C8\_RAT STANDARD; PRT; 482 AA.  
AC Q10758; Q5WBP3;  
DT 01-OCT-1996, integrated into UniProtKB/Swiss-Prot.



RL Gene 200:149-156(1997).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Suzuki Y., Sugano S., Toroki Y., Toyoda A., Takeda T., Sakaki Y.,  
RA Tanaka A., Yokoyama S.,  
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK222941; BAD96661.1; -; mRNA.  
DR Ensembl; ENSG00000170421; Homo sapiens.  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR InterPro; IPR003054; Keratin\_II.  
DR PANTHER; PTHR18893:SF5; Keratin\_II; 1.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
FT NON\_TER 1  
SQ SEQUENCE 483 AA; 53738 MW; D4F01FEB859C3D3 CRC64;  
  
Query Match 100.0%; Score 43; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. NO. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KYEDEINK 8  
Db 178 KYEDEINK 185  
|||||||

Search completed: July 5, 2006, 19:01:32  
Job time : 83.9641 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 69.5146 Seconds  
(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928A-12  
Perfect score: 43  
Sequence: 1 KYEDELNK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 8:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003Bs:\*  
8: Geneseqp2004s:\*  
9: Geneseqp2005s:\*  
10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	8	6	Aao26392 Psoriasis
2	43	100.0	8	9	ADVO4417 Leishman
3	43	100.0	113	6	ABG74871 Human cyt
4	43	100.0	132	8	ADR98929 Lung spec
5	43	100.0	180	8	ADN99441 Novel hum
6	43	100.0	186	4	Aam16318 Peptide #
7	43	100.0	186	4	ABB35310 Peptide #
8	43	100.0	186	4	Aam28813 Peptide #
9	43	100.0	186	4	ABB30140 Peptide #
10	43	100.0	186	4	ABB20752 Protein #
11	43	100.0	186	4	Aam68514 Human bon
12	43	100.0	186	4	Aam56141 Human bra
13	43	100.0	186	4	ABG50183 Human liv
14	43	100.0	186	4	Aam04056 Peptide #
15	43	100.0	186	4	ABG09410 Novel hum
16	43	100.0	186	5	ABG38095 Human'pep
17	43	100.0	233	8	ADR98930 Lung spec
18	43	100.0	273	4	ABG26376 Novel hum
19	43	100.0	275	4	ABG15282 Novel hum
20	43	100.0	278	8	ABM81219 Tumour-as
21	43	100.0	282	8	ABM80614 Tumour-as
22	43	100.0	342	7	ADE15641 Human str
23	43	100.0	347	4	ABG15636 Novel hum

24	43	100.0	347	4	ABG09411	Abg09411 Novel hum
25	43	100.0	394	4	ABG26375	Abg26375 Novel hum
26	43	100.0	398	6	AAE38086	Aae38086 Human cyt
27	43	100.0	401	8	ABM80365	Abm80365 Tumour-as
28	43	100.0	403	8	ABM81640	Abm81640 Tumour-as
29	43	100.0	420	8	ADP29879	Adp29879 Human sec
30	43	100.0	429	4	ABG19547	Abg19547 Novel hum
31	43	100.0	429	4	ABG19546	Abg19546 Novel hum
32	43	100.0	476	8	ADR66033	Adr66033 Human pro
33	43	100.0	476	8	ADR66931	Adr66931 Human pro
34	43	100.0	477	4	AAU06112	Aau06112 Novel hum
35	43	100.0	480	10	AEF92273	Aef92273 Human cyt
36	43	100.0	481	6	ABU52604	Abu52604 Human NOV
37	43	100.0	482	4	ABG08132	Abg08132 Novel hum
38	43	100.0	482	6	AAE38082	Aae38082 Human cyt
39	43	100.0	482	8	ADT51418	Adt51418 Human ker
40	43	100.0	482	8	ADT51421	Adt51421 Human ker
41	43	100.0	482	8	ADT51419	Adt51419 Human ker
42	43	100.0	482	8	ADT51422	Adt51422 Human ker
43	43	100.0	482	8	ADT51414	Adt51414 Human ker
44	43	100.0	482	8	ADT51420	Adt51420 Human ker
45	43	100.0	482	8	ADT51415	Adt51415 Human ker

## ALIGNMENTS

RESULT 1  
AAO26392  
ID AAO26392 standard; peptide; 8 AA.  
XX  
AC AAO26392;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 12.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
PA (AKIV-) AKIVA LLC.  
XX  
PI O'daly JA;  
XX  
DR WPI; 2003-018763/01.  
XX  
PT New polypeptide and immunogenic variants comprising amino acid sequences  
PT of particulate antigens, useful for the treatment and clinical remission  
PT of psoriasis.  
XX  
XX Claim 1; Page 43; 56pp; English.  
XX  
CC The invention relates to a polypeptide comprising an isolated amino acid  
CC sequence or immunogenic variants selected from any of 14 fully defined  
CC sequences of 7-16 amino acids, given in the specification. The  
CC immunotherapeutic agents and a pharmaceutical compositions comprising  
CC polynucleotides and vectors of the invention are useful for the treatment  
CC and clinical remission of psoriasis. The isolated nucleic acids are  
CC useful as probes. The sequences of the invention can be used in the  
CC treatment of disorders by gene therapy. This sequence represents one of  
CC the 14 immunotherapeutic peptides of the invention

SQ Sequence 8 AA;

Query Match 100.0%; Score 43; DB 6; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||||

DB 1 KYEDEINK 8

RESULT 2

ADV04417

ID ADV04417 standard; peptide; 8 AA.

XX

AC ADV04417;

XX

DT 24-FEB-2005 (first entry)

XX

DE Leishmania immunotherapeutic peptide SEQ ID NO:12.

XX

KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;

KW psoriasis.

XX

OS Leishmania sp.

XX

PN US2004241168-A1.

XX

PD 02-DEC-2004.

XX

PF 09-FEB-2004; 2004US-00774928.

XX

PR 16-MAR-2001; 2001US-00809003.

PR 17-OCT-2003; 2003US-00687892.

XX

PA (ODAL/) ODALY J A.

XX

PI Odaly JA;

XX

DR WPI; 2005-011563/01.

XX

PT Inhibiting selectively T-cell rolling in human, by administering compound

PT interfering with cutaneous lymphocyte antigen-E selectin interaction and

PT leukocyte function associated antigen-1 interaction.

XX

PS Claim 18; SEQ ID NO 12; 21pp; English.

XX

CC The invention relates to a novel method for selectively inhibiting T-cell

CC rolling in a human host, comprising administering a compound that

CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E

CC selectin interaction and leukocyte function associated antigen (LFA)-

CC 1/intercellular adhesion molecule (ICAM) and very late antigen

CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound

CC includes an immunotherapeutic agent, which comprises a purified protein

CC extract that is isolated by diethylaminoethyl Sephadex chromatography of

CC a Nonidet P-40 insoluble particulate antigen fraction derived from

CC isolated killed cells of anastigotes from one or more species of the

CC Leishmania genus, where the particulate antigen fraction is solubilized

CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied

CC to diethylaminoethyl Sephadex and is eluted with a solution comprising

CC 0.1 M sodium chloride, 8 M urea and 0.025 M

CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract

CC includes polypeptides having apparent molecular weights of 73, 80 and 82

CC kDa, after total reduction and alkylation. The species is L. amazonensis,

CC L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the

CC invention has antipsoriatic and immunostimulant activity, and inhibits T-

CC cell rolling by interfering with CLA-E selectin interaction and LFA-

CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively

CC inhibiting T-cell rolling in a human host, and for treating psoriasis.

CC The present sequence represents a Leishmania peptide used in the

CC invention.

XX

Sequence 8 AA;

Query Match 100.0%; Score 43; DB 9; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.1e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||||

DB 1 KYEDEINK 8

RESULT 3

ABG74871

ID ABG74871 standard; protein; 113 AA.

XX

AC ABG74871;

XX

DT 19-JUN-2003 (first entry)

XX

DE Human cytokeraatin-1 protein fragment.

XX

KW Cytokeratin-1; inflammation; infection; soluble; sepsis; veterinary;

KW sepsis-like systemic infection; human.

XX

OS Homo sapiens.

XX

PN WO2003002600-A1.

XX

PD 09-JAN-2003.

XX

PF 12-JUN-2002; 2002WO-EP006473.

XX

PR 27-JUN-2001; 2001DE-01030985.

XX

PA (BRAH-) BRAHMS AG.

XX

PI Bergmann A, Struck J, Uehlein M;

XX

DR WPI; 2003-201491/19.

XX

PT New use of soluble cytokeraatin-1 fragments from body fluid are tissue as

PT marker peptides or to provide therapeutics for the diagnosis, prognosis

PT and treatment of inflammatory disease and infection, including systemic

PT inflammation.

XX

PS Claim 4; Page 41; Slipp; German.

XX

CC This invention describes a novel use of soluble cytokeraatin-1 fragments

CC from body fluid or tissue as marker peptides for the detection, prognosis

CC and control of inflammation and infection or as a therapeutic target for

CC inflammation and infection. The invention also describes a method of

CC producing a soluble cytokeraatin-1 fragment in rich or pure form from a

CC human material rich in cytokeraatin-1 by enzymatically splitting the

CC cytokeraatin-1 with endoprotease to obtain the hydrolysis products and

CC purifying the resulting fragments. The novel cytokeraatin fragments are

CC useful for the differential diagnosis, early recognition and prognosis,

CC evaluation of severity of sepsis and heavy infection, particularly sepsis

CC -like systemic infection. The fragments are used in human and veterinary

CC therapy, particularly the fragment or it's specific antibody is used to

CC treat inflammatory disease or infection. This sequence represents the

CC human cytokeraatin-1 protein fragment described in the disclosure of the

CC invention

XX

Sequence 113 AA;

Query Match 100.0%; Score 43; DB 6; Length 113;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||||

DB 84 KYEDEINK 91



```
RESULT 4
ADR98929
ID ADR98929 standard; protein; 132 AA.
XX AC ADR98929;
XX DT 18-NOV-2004 (first entry)
XX DE Lung specific gene splice variant encoded protein #101.
XX KW cytostatic; gene therapy; vaccine; lung; diagnosis; cancer;
XX KW non-cancerous lung disease; lung tissue; antagonist; gene therapy;
XX KW transgenic animal; splice variant.
XX OS Homo sapiens.
XX PN WO2004074430-A2.
XX PD 02-SEP-2004.
XX PF 08-DEC-2003; 2003WO-US038896.
XX PR 06-DEC-2002; 2002US-0431307P.
XX PR 06-DEC-2002; 2002US-0431510P.
XX PR 06-DEC-2002; 2002US-0431516P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Macina RA, Turner LR, Sun Y, Liu S;
XX WPI: 2004-635553/61.
XX N-PSDB; ADR98791.
XX PT New isolated human lung specific nucleic acid molecule, useful for
XX PT identifying, diagnosing, monitoring, staging, imaging and treating lung
XX PT cancer and non-cancerous diseases of the lung.
XX PS Claim 1; SEQ ID NO 214; 542pp; English.
XX CC The invention relates to a new isolated lung specific nucleic acid
XX CC molecule (1) comprising any of 113 fully defined nucleotide sequences
XX CC given in the specification, their encoded protein sequences, sequences
XX CC selectively hybridizing to the nucleotide sequences or a sequence having
XX CC at least 60% identity to the nucleotide sequences. The methods and
XX CC compositions of the present invention are useful for identifying,
XX CC diagnosing, monitoring, staging, imaging and treating lung cancer and non
XX CC -cancerous diseases of the lung. They are also used for identifying lung
XX CC tissue, monitoring and identifying and/or designing antagonists of the
XX CC polypeptide of the invention, gene therapy, production of transgenic
XX CC animals and production of engineered lung tissue for treatment and
XX CC research. Lung specific genes (LSGs) were identified by a systematic
XX CC analysis of gene expression data in the LIFESeq Gold database using the
XX CC data mining software package candidate lead automatic search program
XX CC (CLASP). Genes were grouped into gene bins where each bin is a cluster of
XX CC sequences grouped together where they share a common contig.
XX CC Differentially expressed tissue-specific genes were selected based on the
XX CC percentage level in the targeted tissue versus all the other tissues. The
XX CC expression levels for each gene in libraries of normal tissues or non-
XX CC tumour tissues from cancer patients were compared with the expression
XX CC levels in tissue libraries associated with tumour or disease. This
XX CC sequence represents a protein of the invention.
XX SQ Sequence 132 AA;
Query Match 100.0%; Score 43; DB 8; Length 132;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KYEIDEINK 8
DB 46 KYEIDEINK 53
```

```
RESULT 5
ADR99441
ID ADR99441 standard; protein; 180 AA.
XX AC ADR99441;
XX DT 29-JUL-2004 (first entry)
XX DE Novel human protein sequence #257.
XX KW anti-inflammatory; dermatological; neuroprotective; immunomodulator;
XX KW antibacterial; virucide; antipsoriatic; cytostatic; gene therapy;
XX KW vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
XX KW early aging; hormonal imbalance; ischemic heart disease;
XX KW ulcerative colitis.
XX OS Homo sapiens.
XX PN WO2004038003-A2.
XX PD 06-MAY-2004.
XX PF 24-OCT-2003; 2003WO-US033947.
XX PR 25-OCT-2002; 2002US-0421061P.
XX PR 25-OCT-2002; 2002US-0421080P.
XX PR 25-OCT-2002; 2002US-0421552P.
XX PR 25-OCT-2002; 2002US-0421614P.
XX PR 30-OCT-2002; 2002US-0422177P.
XX PR 30-OCT-2002; 2002US-0422178P.
XX PR 15-NOV-2002; 2002US-0426355P.
XX PR 15-NOV-2002; 2002US-0426384P.
XX PR 15-NOV-2002; 2002US-0426394P.
XX PR 15-NOV-2002; 2002US-0426430P.
XX PR 15-NOV-2002; 2002US-0426916P.
XX PR 27-NOV-2002; 2002US-0429224P.
XX PR 27-NOV-2002; 2002US-0429275P.
XX PR 27-NOV-2002; 2002US-0429302P.
XX PR 27-NOV-2002; 2002US-0429326P.
XX PR 27-NOV-2002; 2002US-0429651P.
XX PR 04-DEC-2002; 2002US-0430645P.
XX PR 04-DEC-2002; 2002US-0430651P.
XX PR 04-DEC-2002; 2002US-0430657P.
XX PR 04-DEC-2002; 2002US-0430663P.
XX PR 04-DEC-2002; 2002US-0430668P.
XX PR 04-DEC-2002; 2002US-0430684P.
XX PR 05-DEC-2002; 2002US-0430937P.
XX PR 05-DEC-2002; 2002US-0430965P.
XX PR 05-DEC-2002; 2002US-0431458P.
XX PR 12-DEC-2002; 2002US-0433251P.
XX PR 12-DEC-2002; 2002US-0433500P.
XX PR 13-DEC-2002; 2002US-0433316P.
XX PR 13-DEC-2002; 2002US-0433318P.
XX PR 23-DEC-2002; 2002US-0436236P.
XX PR 03-JAN-2003; 2003US-0437914P.
XX PR 17-JAN-2003; 2003US-0440820P.
XX PR 17-JAN-2003; 2003US-0440821P.
XX PR 18-APR-2003; 2003US-0463700P.
XX PR 18-APR-2003; 2003US-0463708P.
XX PR 18-APR-2003; 2003US-0463716P.
XX PR 18-APR-2003; 2003US-0463732P.
XX PR 02-MAY-2003; 2003US-0467199P.
XX PR 02-MAY-2003; 2003US-0467201P.
XX PR 02-MAY-2003; 2003US-0467203P.
XX PR 02-MAY-2003; 2003US-0467230P.
XX PR 19-MAY-2003; 2003US-0471306P.
XX PR 19-MAY-2003; 2003US-0471336P.
XX PR 22-MAY-2003; 2003US-0472420P.
XX PR 22-MAY-2003; 2003US-0472430P.
XX PR 09-JUN-2003; 2003US-0476609P.
XX PR 09-JUN-2003; 2003US-0476621P.
XX PR 09-JUN-2003; 2003US-0476632P.
XX PR 09-JUN-2003; 2003US-0476641P.
```

PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
XX	(MOLE-) MOLECULAR DYNAMICS INC.
PA	
XX	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI; 2001-488901/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for analyzing
PT	gene expression in human cervical epithelial cells.
XX	
PS	Claim 27; SEQ ID NO 21144; 487pp; English.
XX	
CC	The present invention relates to human single exon nucleic acid probes
CC	(SENPs; see AAI10066-AAI28459). The present sequence is a peptide encoded
CC	by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC	can be used to produce a single exon microarray, which can be used for
CC	measuring human gene expression in a sample derived from human cervical
CC	epithelial cells. By measuring gene expression, the probes are therefore
CC	useful in grading and/or staging of diseases of the cervix, notably
CC	cervical cancer. Note: The sequence data for this patent did not form
CC	part of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
XX	Sequence 186 AA;
SQ	
	Query Match 100.0%; Score 43; DB 4; Length 186;
	Best Local Similarity 100.0%; Pred. No. 4.6;
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Qy	1 KYEDEINK 8
Db	173 KYEDEINK 180
RESULT 7	
ABB35310	
ID	ABB35310 standard; peptide; 186 AA.
XX	AC ABB35310;
XX	
DT	04-FEB-2002 (first entry)
XX	
DE	Peptide #2816 encoded by human foetal liver single exon probe.
XX	
KW	Human; foetal liver; gene expression; single exon nucleic acid probe.
XX	
OS	Homo sapiens.
XX	
PN	WO200157277-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US000669.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-483447/52.
XX	

PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human fetal liver.  
PS Claim 27; SEQ ID NO 27945; 639pp + Sequence Listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human fetal liver. The  
CC present sequence is a peptide encoded by a single exon nucleic acid probe  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 186 AA;

Query Match 100.0%; Score 43; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
Db 173 KYEDEINK 180  
|||||

RESULT 8  
AAM28813  
ID AAM28813 standard; protein; 186 AA.  
XX  
AC AAM28813;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Peptide #2850 encoded by probe for measuring placental gene expression.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000663.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488897/53.  
XX  
PF 30-JAN-2001; 2001WO-US000663.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488897/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX  
PS Claim 27; SEQ ID NO 29082; 654pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENP:  
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX  
SQ Sequence 186 AA;

Query Match 100.0%; Score 43; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
Db 173 KYEDEINK 180  
|||||

RESULT 8  
AAM28813  
ID AAM28813 standard; protein; 186 AA.  
XX  
AC AAM28813;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Peptide #2850 encoded by probe for measuring placental gene expression.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000663.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488897/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX  
PS Claim 27; SEQ ID NO 29082; 654pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENP:  
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX  
SQ Sequence 186 AA;

Query Match 100.0%; Score 43; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
Db 173 KYEDEINK 180  
|||||

RESULT 9  
ABB30140  
ID ABB30140 standard; peptide; 186 AA.  
XX  
AC ABB30140;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Peptide #2791 encoded by breast cell single exon nucleic acid probe.  
XX  
KW Human; microarray; single exon probe; gene expression; breast; disease;  
KW cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157271-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000662.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-496933/54.  
XX  
PT New spatially-addressable set of single exon nucleic acid probes, useful  
PT for measuring gene expression in sample derived from human breast,  
PT comprises number of single exon nucleic acid probes.  
XX  
PS Claim 27; SEQ ID NO 13108; 327pp + Sequence Listing; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting the  
CC probes with a collection of detectably labelled nucleic acids derived  
CC from mRNA of human breast, and then measuring the label bound to each  
CC probe of the microarray. The probes are useful for verifying the  
CC expression of regions of genomic DNA predicted to encode proteins. They  
CC are useful for gene discovery, and for determining predisposition and/or  
CC prognosing breast disease. Gene expression analysis is useful for  
CC assessing the toxicity of chemical agents on cells. The microarray of  
CC this invention presents a far greater diversity of probes for measuring  
CC gene expression, with far less bias than expressed sequence tag  
CC microarrays. The method is suitable for rapid production of functional  
CC information from genomic sequence. The present sequence is a peptide  
CC encoded by a single exon nucleic acid probe of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 186 AA;

Query Match 100.0%; Score 43; DB 4; Length 186;

```
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 173 KYEDEINK 180

RESULT 10
ABB20752
ID ABB20752 standard; protein; 186 AA.
XX AC ABB20752;
XX DT 23-JAN-2002 (first entry)
XX DE Protein #2751 encoded by probe for measuring heart cell gene expression.
XX KW Human; gene expression; heart; microarray; vascular system;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX XX 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488999/53.
XX XX Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 15; SEQ ID NO 22522; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart (see
XX CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX CC probe. The probes may be used for predicting, measuring and displaying
XX CC gene expression in samples derived from the human heart via microarrays.
XX CC By measuring gene expression, the probes are useful for predicting,
XX CC diagnosing, grading, staging, monitoring and prognosing diseases of the
XX CC human heart and vascular system e.g. cardiovascular disease,
XX CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 186 AA;

Query Match 100.0%; Score 43; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 173 KYEDEINK 180
```

```
RESULT 11
AAM68514
ID AAM68514 standard; protein; 186 AA.
XX AC AAM68514;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28820.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX XX 09-AUG-2001.
XX PD 30-JAN-2001; 2001WO-US000668.
XX PF 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488999/53.
XX XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO 28820; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention
XX SQ Sequence 186 AA;

Query Match 100.0%; Score 43; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 173 KYEDEINK 180

RESULT 12
AAM56141
ID AAM56141 standard; protein; 186 AA.
XX AC AAM56141;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 28246.
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX XX
```

```
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 28246; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
XX SQ Sequence 186 AA;
XX
XX Query Match 100.0%; Score 43; DB 4; Length 186;
XX Best Local Similarity 100.0%; Pred. No. 4.6;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 KYEDEINK 8
XX Db 173 KYEDEINK 180
XX
XX RESULT 13
XX ABG50183
XX ID ABG50183 standard; peptide; 186 AA.
XX
XX AC ABG50183;
XX
XX DT 25-FEB-2003 (first entry)
XX
XX DE Human liver peptide, SEQ ID No 28831.
XX
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX
XX OS Homo sapiens.
XX
XX PN WO200157273-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX
```

---

```
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 28831; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 186 AA;
XX
XX Query Match 100.0%; Score 43; DB 4; Length 186;
XX Best Local Similarity 100.0%; Pred. No. 4.6;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 KYEDEINK 8
XX Db 173 KYEDEINK 180
XX
XX RESULT 14
XX AAM04056
XX ID AAM04056 standard; protein; 186 AA.
XX
XX AC AAM04056;
XX
XX DT 09-OCT-2001 (first entry)
XX
XX DE Peptide #2738 encoded by probe for measuring breast gene expression.
XX
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX OS Homo sapiens.
XX
XX PN WO200157270-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 29-JAN-2001; 2001WO-US000661.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX
```

PT a human breast.  
PS Claim 27; SEQ ID NO 12796; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes  
CC (see AA100010-AA110067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours. Note: The sequence data for  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 186 AA;  
Query Match 100.0%; Score 43; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 KYEDEINK 8  
Db 173 KYEDEINK 180  
RESULT 15  
ABG09410  
ID ABG09410 standard; protein; 186 AA.  
XX  
AC ABG09410;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #9401.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS73597.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 39769; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 186 AA;

Query Match 100.0%; Score 43; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8  
Db 11 KYEDEINK 18

Search completed: July 5, 2006, 18:43:52  
Job time : 70.5146 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 58.5631 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928A-12

Perfect score: 43

Sequence: 1 KYEDEINK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	43	100.0	8	5	US-10-774-928-12	Sequence 12, Appl
2	43	100.0	113	5	US-10-480-848A-5	Sequence 5, Appli
3	43	100.0	186	3	US-09-864-761-36050	Sequence 36050, A
4	43	100.0	186	3	US-10-450-763-39769	Sequence 39769, A
5	43	100.0	273	5	US-10-450-763-56735	Sequence 56735, A
6	43	100.0	275	5	US-10-450-763-45641	Sequence 45641, A
7	43	100.0	347	5	US-10-450-763-39770	Sequence 39770, A
8	43	100.0	347	5	US-10-450-763-45995	Sequence 45995, A
9	43	100.0	384	5	US-10-450-763-56734	Sequence 56734, A
10	43	100.0	398	5	US-10-883-020-5	Sequence 5, Appli
11	43	100.0	422	3	US-09-779-307-18	Sequence 18, Appl
12	43	100.0	422	3	US-09-779-307-20	Sequence 20, Appl
13	43	100.0	429	5	US-10-450-763-49905	Sequence 49905, A
14	43	100.0	429	5	US-10-450-763-49906	Sequence 49906, A
15	43	100.0	430	3	US-09-779-307-19	Sequence 19, Appl
16	43	100.0	476	3	US-09-779-307-16	Sequence 16, Appl
17	43	100.0	477	3	US-09-779-307-4	Sequence 4, Appli
18	43	100.0	481	4	US-10-080-334-54	Sequence 54, Appl
19	43	100.0	481	6	US-11-156-300-24	Sequence 24, Appl
20	43	100.0	482	3	US-09-779-307-17	Sequence 17, Appl
21	43	100.0	482	4	US-10-080-334-187	Sequence 187, App
22	43	100.0	482	5	US-10-883-020-1	Sequence 1, Appli
23	43	100.0	482	5	US-10-450-763-38491	Sequence 38491, A
24	43	100.0	483	3	US-09-919-497-79	Sequence 79, Appl
25	43	100.0	483	3	US-09-974-298-41	Sequence 41, Appl
26	43	100.0	483	3	US-09-981-353-158	Sequence 158, App
27	43	100.0	483	4	US-10-177-293-258	Sequence 258, App

28 43 100.0 483 4 US-10-080-334-183 Sequence 183, App  
29 43 100.0 483 4 US-10-080-334-184 Sequence 184, App  
30 43 100.0 483 4 US-10-080-334-185 Sequence 185, App  
31 43 100.0 483 4 US-10-080-334-186 Sequence 186, App  
32 43 100.0 483 4 US-10-734-564-116 Sequence 116, App  
33 43 100.0 483 6 US-11-037-713-9 Sequence 9, Appli  
34 43 100.0 498 4 US-10-605-708A-2 Sequence 2, Appli  
35 43 100.0 500 5 US-10-450-763-39773 Sequence 39773, A  
36 43 100.0 513 5 US-10-450-763-38492 Sequence 38492, A  
37 43 100.0 521 3 US-09-976-782-8 Sequence 8, Appli  
38 43 100.0 523 3 US-09-976-782-6 Sequence 6, Appli  
39 43 100.0 541 5 US-10-450-763-55674 Sequence 55674, A  
40 43 100.0 564 4 US-10-170-385-421 Sequence 421, App  
41 43 100.0 564 5 US-10-723-860-1619 Sequence 1619, App  
42 43 100.0 564 5 US-10-756-149-5130 Sequence 5130, App  
43 43 100.0 564 5 US-10-631-467-938 Sequence 938, App  
44 43 100.0 564 6 US-11-037-713-27 Sequence 27, Appl  
45 43 100.0 567 4 US-10-104-047-3772 Sequence 3772, App

ALIGNMENTS

RESULT 1

US-10-774-928-12  
; Sequence 12, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-12

Query Match 100.0%; Score 43; DB 5; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.9e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||  
DB 1 KYEDEINK 8

RESULT 2

US-10-480-848A-5  
; Sequence 5, Application US/10480848A  
; Publication No. US20040219597A1  
; GENERAL INFORMATION:  
; APPLICANT: B.R.A.H.M.S Aktiengesellschaft  
; TITLE OF INVENTION: Use of soluble cytokeatin 1 fragments in diagnosis and  
; TITLE OF INVENTION: therapy  
; FILE REFERENCE: 3537PCTNAT  
; CURRENT APPLICATION NUMBER: US/10/480,848A  
; CURRENT FILING DATE: 2003-12-15  
; PRIOR APPLICATION NUMBER: DE 101 30 985.6  
; PRIOR FILING DATE: 2001-06-27  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-480-848A-5

Query Match 100.0%; Score 43; DB 5; Length 113;

```
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
Db 84 KYEDEINK 91

RESULT 3
US-09-864-761-36050
; Sequence 36050, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-09-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36050
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011309.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 72
; OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 59
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.8
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 44
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5
```

```
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5e+02
; OTHER INFORMATION: EST HUMAN HIT: BE870853.1, EVALUE 9.00e-68
; OTHER INFORMATION: SWISSPROT HIT: P05787, EVALUE 1.00e-68
US-09-864-761-36050

Query Match 100.0%; Score 43; DB 3; Length 186;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
Db 173 KYEDEINK 180

RESULT 4
US-10-450-763-39769
; Sequence 39769, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 39769
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (124)..(171)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226D, p-value=7.000e-38, raw score of 19.10
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(169)
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,
; OTHER INFORMATION: accession name filament, E-value=4.3e-71, Pfam score of 249.6
US-10-450-763-39769

Query Match 100.0%; Score 43; DB 5; Length 186;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8
Db 11 KYEDEINK 18

RESULT 5
US-10-450-763-56735
; Sequence 56735, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
```



;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO 56735  
;; LENGTH: 273  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (177)..(225)  
;; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,  
;; accession number BL00226B, p-value=3.348e-35, raw score of 23.86  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (78)..(271)  
;; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
;; accession number BL00226B, p-value=8.6e-101, Pfam score of 348.3  
US-10-450-763-56735

Query Match 100.0%; Score 43; DB 5; Length 273;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
DB 166 KYEDEINK 173

## RESULT 6

US-10-450-763-45641  
;; Sequence 45641, Application US/10450763  
;; Publication No. US20050196754A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
;; FILE REFERENCE: 790CIP3/US  
;; CURRENT APPLICATION NUMBER: US/10/450,763  
;; CURRENT FILING DATE: 2003-06-11  
;; PRIOR APPLICATION NUMBER: PCT/US01/08631  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 09/540,217  
;; PRIOR FILING DATE: 2000-03-31  
;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO 45641  
;; LENGTH: 275  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (179)..(227)  
;; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,  
;; accession number BL00226B, p-value=3.348e-35, raw score of 23.86  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (87)..(273)  
;; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
;; accession number BL00226B, p-value=3.2e-89, Pfam score of 309.8  
US-10-450-763-45641

Query Match 100.0%; Score 43; DB 5; Length 275;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
DB 168 KYEDEINK 175

## RESULT 7

US-10-450-763-39770  
;; Sequence 39770, Application US/10450763  
;; Publication No. US20050196754A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
;; FILE REFERENCE: 790CIP3/US  
;; CURRENT APPLICATION NUMBER: US/10/450,763  
;; CURRENT FILING DATE: 2003-06-11  
;; PRIOR APPLICATION NUMBER: PCT/US01/08631  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 09/540,217  
;; PRIOR FILING DATE: 2000-03-31  
;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO 39770  
;; LENGTH: 347  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (180)..(228)  
;; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,  
;; accession number BL00226B, p-value=1.643e-31, raw score of 23.86  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (77)..(342)  
;; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
;; accession number BL00226B, p-value=1.1e-78, Pfam score of 274.8  
US-10-450-763-39770

Query Match 100.0%; Score 43; DB 5; Length 347;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
DB 169 KYEDEINK 176

## RESULT 8

US-10-450-763-45995  
;; Sequence 45995, Application US/10450763  
;; Publication No. US20050196754A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
;; FILE REFERENCE: 790CIP3/US  
;; CURRENT APPLICATION NUMBER: US/10/450,763  
;; CURRENT FILING DATE: 2003-06-11  
;; PRIOR APPLICATION NUMBER: PCT/US01/08631  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 09/540,217  
;; PRIOR FILING DATE: 2000-03-31  
;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO 45995  
;; LENGTH: 347  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (180)..(228)  
;; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,  
;; accession number BL00226B, p-value=1.643e-31, raw score of 23.86  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (77)..(342)  
;; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
;; accession number BL00226B, p-value=1.1e-78, Pfam score of 274.8  
US-10-450-763-45995

Query Match 100.0%; Score 43; DB 5; Length 275;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
DB 168 KYEDEINK 175

## RESULT 7

```
; OTHER INFORMATION: accession name filament, E-value=1.1e-78, Pfam score of 274.8
US-10-450-763-45995

Query Match      100.0%; Score 43; DB 5; Length 347;
Best Local Similarity 100.0%; Pred. No. 16;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      169 KYEDEINK 176

RESULT 9
US-10-450-763-56734
; Sequence 56734, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 56734
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (164)..(212)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00226B, p-value=4.656e-29, raw score of 23.86
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (80)..(312)
; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,
; OTHER INFORMATION: accession name filament, E-value=4e-88, Pfam score of 306.2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(394)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-56734

Query Match      100.0%; Score 43; DB 5; Length 394;
Best Local Similarity 100.0%; Pred. No. 18;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      153 KYEDEINK 160

RESULT 10
US-10-883-020-5
; Sequence 5, Application US/10883020
; Publication No. US20050048070A1
; GENERAL INFORMATION:
; APPLICANT: Ditzel, Henrik
; APPLICANT: Jensenius, Jens
; TITLE OF INVENTION: The Scripps Research Institute
; FILE REFERENCE: 1361.017US1
; CURRENT APPLICATION NUMBER: US/10/883,020
; CURRENT FILING DATE: 2004-07-01
; PRIOR APPLICATION NUMBER: PCT/US03/00297
; PRIOR FILING DATE: 2003-01-03
```

```
; PRIOR APPLICATION NUMBER: US 60/345,208
; PRIOR FILING DATE: 2002-01-03
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-883-020-5

Query Match      100.0%; Score 43; DB 5; Length 398;
Best Local Similarity 100.0%; Pred. No. 18;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      93 KYEDEINK 100

RESULT 11
US-09-779-307-18
; Sequence 18, Application US/09779307
; Patent No. US20020137675A1
; GENERAL INFORMATION:
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Prayaga, Sudhirdas
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-662 US
; CURRENT APPLICATION NUMBER: US/09/779,307
; CURRENT FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/180,880
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/181,044
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/181,656
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/182,795
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-779-307-18

Query Match      100.0%; Score 43; DB 3; Length 422;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      177 KYEDEINK 184

RESULT 12
US-09-779-307-20
; Sequence 20, Application US/09779307
; Patent No. US20020137675A1
; GENERAL INFORMATION:
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Prayaga, Sudhirdas
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-662 US
; CURRENT APPLICATION NUMBER: US/09/779,307
; CURRENT FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/180,880
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/181,044
```

; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 60/181,656  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: 60/182,795  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 422  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-779-307-20

Query Match 100.0%; Score 43; DB 3; Length 422;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||  
Db 177 KYEDEINK 184

RESULT 13  
US-10-450-763-49905  
; Sequence 49905, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 49905  
; LENGTH: 429  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (183)..(231)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by eMATRIX.  
; OTHER INFORMATION: accession number BL00226B, p-value=2.714e-23, raw score of 23.86  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (107)..(348)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by PFam,  
; OTHER INFORMATION: accession name filament, E-value=4.2e-90, PFam score of 312.7  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(429)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-49905

Query Match 100.0%; Score 43; DB 5; Length 429;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||  
Db 172 KYEDEINK 179

RESULT 14  
US-10-450-763-49906  
; Sequence 49906, Application US/10450763  
; Publication No. US20050196754A1

; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 49906  
; LENGTH: 429  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (183)..(231)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by eMATRIX,  
; OTHER INFORMATION: accession number BL00226B, p-value=2.714e-23, raw score of 23.86  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (107)..(348)  
; OTHER INFORMATION: Intermediate filament proteins domain identified by PFam,  
; OTHER INFORMATION: accession name filament, E-value=4.2e-90, PFam score of 312.7  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(429)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-49906

Query Match 100.0%; Score 43; DB 5; Length 429;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
|||  
Db 172 KYEDEINK 179

RESULT 15  
US-09-779-307-19  
; Sequence 19, Application US/09779307  
; Patent No. US20020137675A1  
; GENERAL INFORMATION:  
; APPLICANT: Taupier Jr., Raymond  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Vernet, Corine  
; APPLICANT: Prayaga, Sudhirdas  
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby  
; FILE REFERENCE: 15966-662 US  
; CURRENT APPLICATION NUMBER: US/09/779,307  
; CURRENT FILING DATE: 2001-02-07  
; PRIOR APPLICATION NUMBER: 60/180,880  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 60/181,044  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 60/181,656  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: 60/182,795  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-779-307-19

Query Match 100.0%; Score 43; DB 3; Length 430;

Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8  
|||  
Db 184 KYEDEINK 191

Search completed: July 5, 2006, 20:51:59  
Job time : 58.5631 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 4.58252 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-12

Perfect score: 43

Sequence: 1 KYEDEINK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.New.\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	483	7	US-11-105-233-180
2	40	93.0	469	6	US-10-505-928-457
3	39	90.7	140	6	US-10-953-349-1458
4	39	90.7	147	6	US-10-953-349-1457
5	39	90.7	164	6	US-10-953-349-1456
6	34	79.1	940	6	US-10-449-902-45238
7	34	79.1	940	6	US-10-449-902-45669
8	33	76.7	242	6	US-10-449-902-46212
9	33	76.7	256	7	US-11-174-307B-4514
10	33	76.7	456	6	US-10-706-435A-26
11	33	76.7	457	6	US-10-706-435A-4
12	33	76.7	758	6	US-10-449-902-51086
13	33	76.7	799	6	US-10-449-902-47602
14	33	76.7	964	7	US-11-289-102-392
15	32	74.4	33	7	US-11-134-871-915
16	32	74.4	163	6	US-10-449-902-38644
17	32	74.4	225	6	US-10-471-571A-5244
18	32	74.4	251	6	US-10-449-902-47327
19	32	74.4	275	6	US-10-449-902-48344
20	32	74.4	436	6	US-10-471-571A-5632
21	32	74.4	691	6	US-10-953-349-1322
22	32	74.4	705	6	US-10-953-349-1321
23	32	74.4	713	6	US-10-953-349-1320
24	32	74.4	723	6	US-10-449-902-53379
25	31	72.1	44	7	US-11-199-489A-136

26 31 72.1 107 6 US-10-953-349-13561 Sequence 13561, A  
27 31 72.1 158 6 US-10-953-349-13560 Sequence 13560, A  
28 31 72.1 196 6 US-10-953-349-19059 Sequence 19059, A  
29 31 72.1 238 6 US-10-953-349-19058 Sequence 19058, A  
30 31 72.1 247 6 US-10-449-902-54216 Sequence 54216, A  
31 31 72.1 273 7 US-11-101-316-168 Sequence 168, App  
32 31 72.1 568 6 US-10-449-902-46926 Sequence 46926, A  
33 31 72.1 568 6 US-10-449-902-47019 Sequence 47019, A  
34 31 72.1 607 6 US-10-449-902-45927 Sequence 45927, A  
35 31 72.1 714 6 US-10-449-902-54951 Sequence 54951, A  
36 31 72.1 2871 6 US-10-505-928-100 Sequence 100, App  
37 30 69.8 228 6 US-10-449-902-50750 Sequence 50750, A  
38 30 69.8 247 6 US-10-953-349-17634 Sequence 17634, A  
39 30 69.8 251 6 US-10-449-902-30710 Sequence 30710, A  
40 30 69.8 251 6 US-10-449-902-32230 Sequence 32230, A  
41 30 69.8 257 6 US-10-953-349-4598 Sequence 4598, Ap  
42 30 69.8 263 6 US-10-953-349-33086 Sequence 33086, A  
43 30 69.8 283 6 US-10-953-349-33085 Sequence 33085, A  
44 30 69.8 286 6 US-10-953-349-4597 Sequence 4597, Ap  
45 30 69.8 304 6 US-10-953-349-33084 Sequence 33084, A

#### ALIGNMENTS

RESULT 1  
US-11-105-233-180  
; Sequence 180, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Thiagalingam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; TITLE OF INVENTION: Tumors  
; FILE REFERENCE: 1657/2001  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; CURRENT FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-105-233-180

Query Match 100.0%; Score 43; DB 7; Length 483;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
DB 178 KYEDEINK 185

RESULT 2  
US-10-505-928-457  
; Sequence 457, Application US/10505928  
; Publication No. US20060088532A1  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig Institute for Cancer Research et al.  
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES  
; FILE REFERENCE: 28967/39178  
; CURRENT APPLICATION NUMBER: US/10/505,928  
; CURRENT FILING DATE: 2004-08-27  
; PRIOR APPLICATION NUMBER: US 60/363,019  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: PatentIn 3.2  
; SEQ ID NO 457  
; LENGTH: 469  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-505-928-457

```
Query Match      93.0%; Score 40; DB 6; Length 469;
Best Local Similarity 87.5%; Pred. No. 2.9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
    |||||:
Db 179 KYEDEINR 186

RESULT 3
US-10-953-349-1458
; Sequence 1458, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1458
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1458

Query Match      90.7%; Score 39; DB 6; Length 140;
Best Local Similarity 87.5%; Pred. No. 1.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
    |||||:
Db 55 KFEDEINK 62

RESULT 4
US-10-953-349-1457
; Sequence 1457, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1457
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1457

Query Match      90.7%; Score 39; DB 6; Length 147;
Best Local Similarity 87.5%; Pred. No. 1.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
    |||||:
Db 62 KFEDEINK 69

RESULT 5
US-10-953-349-1456
; Sequence 1456, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
```

```
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1456
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-1456

Query Match      90.7%; Score 39; DB 6; Length 164;
Best Local Similarity 87.5%; Pred. No. 1.4;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
    |||||:
Db 79 KFEDEINK 86

RESULT 6
US-10-449-902-45238
; Sequence 45238, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45238
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-45238

Query Match      79.1%; Score 34; DB 6; Length 940;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YEDEINK 8
    |||||:
Db 413 YEDEINE 419

RESULT 7
US-10-449-902-45669
; Sequence 45669, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
```

; SEQ ID NO 45669  
; LENGTH: 940  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-45669

Query Match 79.1%; Score 34; DB 6; Length 940;  
Best Local Similarity 85.7%; Pred. No. 83;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YEDEINK 8  
|||||  
Db 413 YEDEINE 419

## RESULT 8

US-10-449-902-46212  
; Sequence 46212, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46212  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-46212

Query Match 76.7%; Score 33; DB 6; Length 242;  
Best Local Similarity 85.7%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YEDEINK 8  
|||||  
Db 84 YEDEANK 90

## RESULT 9

US-11-174-307B-4514  
; Sequence 4514, Application US/11174307B  
; Publication No. US20060143729A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai  
; APPLICANT: BROVER, Vyacheslav  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY  
; FILE OF INVENTION: USEFUL FOR MODIFYING PLANT CHARACTERISTICS  
; FILE REFERENCE: 2750-1601PUS2  
; CURRENT APPLICATION NUMBER: US/11/174,307B  
; CURRENT FILING DATE: 2005-06-30  
; PRIOR APPLICATION NUMBER: 60/583,671  
; PRIOR FILING DATE: 2004-06-30  
; PRIOR APPLICATION NUMBER: 60/583,781  
; PRIOR FILING DATE: 2004-06-30  
; PRIOR APPLICATION NUMBER: 60/583,651  
; PRIOR FILING DATE: 2004-06-30  
; NUMBER OF SEQ ID NOS: 5544  
; SEQ ID NO 4514  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION:  
; OTHER INFORMATION: GI Number: 50947575; NR Description: thylakoid lumen protein,  
; OTHER INFORMATION: chloroplast precursor-like [Oryza sativa (japonica cultivar-group)]  
; OTHER INFORMATION: >gi|42408803|dbj|BAD10064.1| thylakoid lumen protein, chloroplast  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 21537211; NR Description: thylakoid lumen protein,  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 32815945; NR Description: Atig76450 [Arabidopsis  
; OTHER INFORMATION: thaliana] >gi|18203439|sp|Q9S720|THUL1\_ARATH Unknown thylakoid lumen  
; OTHER INFORMATION: protein, chloroplast precursor >gi|18411110|ref|NP\_565131.1|  
; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION:  
; OTHER INFORMATION: GI Number: 61826057; NR Description: PREDICTED: similar to  
; OTHER INFORMATION: serine\arginine repetitive matrix 1, partial [bos taurus]  
; OTHER INFORMATION: gi|61818041|ref|XP\_593433.1| PREDICTED: similar to serine\arginin  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION:  
; OTHER INFORMATION: GI Number: 7949115; NR Description: Ser/Arg-related nuclear  
; OTHER INFORMATION: matrix protein; plenty-of-prolines-101; serine\arginine repetitive  
; OTHER INFORMATION: matrix protein 1 [Mus musculus] >gi|3153821|gb|AAC17422.1|  
US-11-174-307B-4514

Query Match 76.7%; Score 33; DB 7; Length 256;  
Best Local Similarity 85.7%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YEDEINK 8  
|||||  
Db 98 YEDEANK 104

## RESULT 10

US-10-706-435A-26  
; Sequence 26, Application US/10706435A  
; Publication No. US20060088547A1  
; GENERAL INFORMATION:  
; APPLICANT: Walter Reed Army Institute of Research  
; APPLICANT: Lanar, David E.  
; APPLICANT: Hillier, Collette J.  
; APPLICANT: Lyon, Jeffrey A.  
; APPLICANT: Angov, Evelina  
; APPLICANT: Kumar, Sanjai  
; APPLICANT: Rogers, William  
; APPLICANT: Barbosa, Arnaldo  
; TITLE OF INVENTION: Expression, Purification, and Uses of a Plasmodium  
; FILE OF INVENTION: falciparum Liver Stage Antigen 1 Polypeptide  
; FILE REFERENCE: 003/285/SAP  
; CURRENT APPLICATION NUMBER: US/10/706,435A  
; CURRENT FILING DATE: 2003-11-12  
; PRIOR APPLICATION NUMBER: 60/425,719  
; PRIOR FILING DATE: 2002-11-12  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: Microsoft Word XP  
; SEQ ID NO 26  
; LENGTH: 456  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: LSA-NRC(H) protein  
US-10-706-435A-26

Query Match 76.7%; Score 33; DB 6; Length 456;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEIN 7

```
Db          294 KYEDELIS 300
|||||:
RESULT 11
US-10-706-435A-4
; Sequence 4, Application US/10706435A
; Publication No. US20060088547A1
; GENERAL INFORMATION:
; APPLICANT: Walter Reed Army Institute of Research
; APPLICANT: Lanar, David E.
; APPLICANT: Hillier, Collette J.
; APPLICANT: Lyon, Jeffrey A.
; APPLICANT: Angov, Evelina
; APPLICANT: Kumar, Sanjai
; APPLICANT: Rogers, William
; APPLICANT: Barbosa, Arnaldo
; TITLE OF INVENTION: Expression, Purification, and Uses of a Plasmodium
; FILE REFERENCE: 003/285/SAP
; CURRENT APPLICATION NUMBER: US/10/706,435A
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: 60/425,719
; PRIOR FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Word XP
; SEQ ID NO 4
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: LSA-NRC (H) Mut
US-10-706-435A-4

Query Match          76.7%; Score 33; DB 6; Length 457;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDELIN 7
|||||:
Db          294 KYEDELIS 300

RESULT 12
US-10-449-902-51086
; Sequence 51086, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51086
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-51086

Query Match          76.7%; Score 33; DB 6; Length 758;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YEDELIN 7
|||||:
Db          230 YEDELIN 235
|||||:
RESULT 13
US-10-449-902-47602
; Sequence 47602, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT CDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47602
; LENGTH: 799
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-47602

Query Match          76.7%; Score 33; DB 6; Length 799;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YEDELIN 7
|||||:
Db          349 YEDELIN 354

RESULT 14
US-11-289-102-392
; Sequence 392, Application US/11289102
; Publication No. US20060121511A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Hyerim
; APPLICANT: Shaw, Peter M.
; APPLICANT: Clark, Edwin
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO
; FILE REFERENCE: 10338 NP
; CURRENT APPLICATION NUMBER: US/11/289,102
; CURRENT FILING DATE: 2005-11-29
; PRIOR APPLICATION NUMBER: US 60/631,993
; PRIOR FILING DATE: 2004-11-30
; NUMBER OF SEQ ID NOS: 395
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 392
; LENGTH: 964
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-289-102-392

Query Match          76.7%; Score 33; DB 7; Length 964;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YEDELIN 8
|||||:
Db          524 YEDEVNQ 530

RESULT 15
US-11-134-871-915
; Sequence 915, Application US/11134871
```



/ Publication No. US20060141528A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Aebersold, Rudolf H.  
/ APPLICANT: Zhang, Hui  
/ TITLE OF INVENTION: Compositions and Methods for  
/ TITLE OF INVENTION: Qualification of Serum Glycoproteins  
/ FILE REFERENCE: 66661-116  
/ CURRENT APPLICATION NUMBER: US/11/134,871  
/ CURRENT FILING DATE: 2005-05-20  
/ PRIOR APPLICATION NUMBER: 60/573,593  
/ PRIOR FILING DATE: 2004-05-21  
/ NUMBER OF SEQ ID NOS: 3602  
/ SOFTWARE: PastSeq for Windows Version 4.0  
/ SEQ ID NO 915  
/ LENGTH: 33  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-11-134-871-915

Query Match 74.4%; Score 32; DB 7; Length 33;  
Best Local Similarity 83.3%; Pred. No. 5.3;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YEDEIN 7  
Db 20 YEDEVN 25

Search completed: July 5, 2006, 20:53:02  
Job time : 4.58252 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 18.0194 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-12  
Perfect score: 43  
Sequence: 1 KYEDEINK 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H\_COMB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/PCTUS\_COMB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE\_COMB.pep:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	8	2	US-09-809-003A-12
2	43	100.0	384	2	US-09-949-016-11034
3	43	100.0	443	2	US-09-949-016-10582
4	43	100.0	482	2	US-09-538-092-858
5	43	100.0	483	2	US-09-919-497-79
6	43	100.0	502	2	US-09-949-016-11033
7	43	100.0	564	2	US-09-949-016-6628
8	43	100.0	567	2	US-10-104-047-3772
9	43	100.0	569	2	US-09-949-016-11035
10	43	100.0	569	2	US-09-949-016-11036
11	43	100.0	637	2	US-09-949-016-8152
12	43	100.0	643	2	US-09-538-092-844
13	43	100.0	645	2	US-09-919-172-41
14	40	93.0	469	2	US-09-077-606-3
15	40	93.0	476	2	US-09-949-016-9096
16	40	93.0	551	1	US-09-067-351-2
17	40	93.0	551	1	US-09-360-490-2
18	39	90.7	608	2	US-10-094-749-2446
19	38	88.4	7	2	US-09-809-003A-1
20	38	88.4	216	2	US-09-248-796A-14967
21	38	88.4	546	1	US-09-067-351-1
22	38	88.4	546	1	US-09-360-490-1
23	34	79.1	234	2	US-09-270-767-32857
24	34	79.1	234	2	US-09-270-767-48074
25	34	79.1	608	2	US-09-270-767-32937
26	34	79.1	608	2	US-09-270-767-48154

27	33	76.7	8	2	US-09-809-003A-11	Sequence 11, Appl
28	33	76.7	229	2	US-09-454-204A-41	Sequence 41, Appl
29	33	76.7	315	2	US-09-248-796A-18213	Sequence 18213, A
30	33	76.7	369	2	US-09-248-796A-18582	Sequence 18582, A
31	33	76.7	407	2	US-09-248-796A-15463	Sequence 15463, A
32	33	76.7	426	2	US-09-538-092-7	Sequence 7, Appl
33	33	76.7	533	2	US-09-248-796A-17526	Sequence 17526, A
34	33	76.7	699	2	US-08-851-843A-52	Sequence 52, Appl
35	33	76.7	699	2	US-08-974-549A-188	Sequence 188, App
36	33	76.7	699	2	US-08-854-050-52	Sequence 52, Appl
37	33	76.7	699	2	US-09-430-323-52	Sequence 52, Appl
38	33	76.7	699	2	US-09-402-181B-188	Sequence 188, App
39	33	76.7	699	2	US-09-721-456-188	Sequence 188, App
40	33	76.7	699	2	US-09-766-253-52	Sequence 52, Appl
41	33	76.7	699	2	US-10-054-295-52	Sequence 52, Appl
42	33	76.7	699	2	US-09-438-486A-52	Sequence 52, Appl
43	33	76.7	699	3	US-10-054-611-52	Sequence 52, Appl
44	33	76.7	719	2	US-08-851-843A-7	Sequence 7, Appl
45	33	76.7	719	2	US-08-974-549A-219	Sequence 219, App

## ALIGNMENTS

RESULT 1  
US-09-809-003A-12  
; Sequence 12, Application US/09809003A  
; Patent No. 6673351  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/09/809,003A  
; CURRENT FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-09-809-003A-12

Query Match 100.0%; Score 43; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 5e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
DB 1 KYEDEINK 8

## RESULT 2

US-09-949-016-11034  
; Sequence 11034, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11034  
; LENGTH: 384

```
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11034

Query Match      100.0%; Score 43; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 275 KYEDEINK 282

RESULT 3
US-09-949-016-10582
; Sequence 10582, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10582
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10582

Query Match      100.0%; Score 43; DB 2; Length 443;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 138 KYEDEINK 145

RESULT 4
US-09-538-092-858
; Sequence 858, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 858
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number P05787
US-09-538-092-858

Query Match      100.0%; Score 43; DB 2; Length 482;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 177 KYEDEINK 184

RESULT 5
US-09-919-497-79
; Sequence 79, Application US/09919497
; Patent No. 6773883
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 79
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-79

Query Match      100.0%; Score 43; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 178 KYEDEINK 185

RESULT 6
US-09-949-016-11033
; Sequence 11033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11033
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11033

Query Match      100.0%; Score 43; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYEDEINK 8
Db 197 KYEDEINK 204
```

```
RESULT 7
US-09-949-016-6628
; Sequence 6628, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6628
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6628

Query Match      100.0%; Score 43; DB 2; Length 564;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      252 KYEDEINK 259

RESULT 8
US-10-104-047-3772
; Sequence 3772, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3772
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3772

Query Match      100.0%; Score 43; DB 2; Length 567;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      234 KYEDEINK 241

RESULT 9
US-09-949-016-11035
; Sequence 11035, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11035
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11035

Query Match      100.0%; Score 43; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      258 KYEDEINK 265

RESULT 10
US-09-949-016-11036
; Sequence 11036, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11036
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11036

Query Match      100.0%; Score 43; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYEDEINK 8
Db      258 KYEDEINK 265

RESULT 11
US-09-949-016-8152
; Sequence 8152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8152
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8152

Query Match      100.0%; Score 43; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1 KYEDEINK 8
DB      269 KYEDEINK 276

RESULT 12
US-09-538-092-844
; Sequence 844, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15366-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 844
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P04264
US-09-538-092-844

Query Match      100.0%; Score 43; DB 2; Length 643;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1 KYEDEINK 8
DB      268 KYEDEINK 275

RESULT 13
US-09-919-172-41
; Sequence 41, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Human
US-09-919-172-41

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673545 1798379CD1
US-09-919-172-41

Query Match      100.0%; Score 43; DB 2; Length 645;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1 KYEDEINK 8
DB      273 KYEDEINK 280

RESULT 14
US-09-077-606-3
; Sequence 3, Application US/09077606
; Patent No. 6774220
; GENERAL INFORMATION:
; APPLICANT: JIANG, Pan Hong
; APPLICANT: KABA, Aboubacar
; APPLICANT: CHANY-FOURNIER, Francoise
; APPLICANT: CERUTTI, Italina
; APPLICANT: CHANY, Charles
; TITLE OF INVENTION: COMPOUNDS HAVING LECTINIC PROPERTIES AND THEIR
; TITLE OF INVENTION: BIOLOGICAL APPLICATIONS
; FILE REFERENCE: 040388/0113
; CURRENT APPLICATION NUMBER: US/09/077,606
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: WO PCT/FR96/01937
; EARLIER FILING DATE: 1996-12-04
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-606-3

Query Match      93.0%; Score 40; DB 2; Length 469;
Best Local Similarity 87.5%; Pred. No. 14;
Matches      7; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

QY      1 KYEDEINK 8
DB      179 KYEDEINK 186

RESULT 15
US-09-949-016-9096
; Sequence 9096, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9096
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9096
```

Query Match 93.0%; Score 40; DB 2; Length 476;  
Best Local Similarity 87.5%; Pred. NO. 14;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYEDEINK 8  
DB 186 KYEDEINR 193

Search completed: July 5, 2006, 19:24:26  
Job time : 19.0194 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:20:41 ; Search time 14.6667 Seconds  
(without alignments)  
104.964 Million cell updates/sec

Title: US-10-774-928a-13

Perfect score: 79

Sequence: 1 EIEQYNLLLASVLD 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*

- 1: PIR1.\*
- 2: PIR2.\*
- 3: PIR3.\*
- 4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	58.2	336	2 T22322	hypothetical prote
2	43	54.4	309	2 T33888	hypothetical prote
3	42	53.2	408	2 T24114	hypothetical prote
4	42	53.2	471	2 E86192	hypothetical prote
5	42	53.2	1613	2 G64488	reverse gyrase (in
6	42	53.2	1624	2 C71129	probable reverse g
7	41	51.9	105	2 E97862	hypothetical prote
8	41	51.9	220	2 S51444	hypothetical prote
9	41	51.9	1092	1 S37676	glutamate dehydrog
10	40	50.6	299	2 F69785	fructokinase homol
11	40	50.6	341	2 S31571	cinnamyl-alcohol d
12	40	50.6	357	2 T09141	cinnamyl-alcohol d
13	40	50.6	358	2 S31572	hypothetical prote
14	40	50.6	771	2 T01315	DNA topoisomerase
15	40	50.6	804	1 ISECTB	hypothetical prote
16	40	50.6	804	2 B91208	DNA gyrase subunit
17	40	50.6	804	2 D86054	hypothetical prote
18	40	50.6	1318	1 Q0BE1	membrane antigen p
19	40	50.6	1318	2 S29606	P140 protein - hum
20	40	50.6	2049	2 T47587	TATA box binding p
21	40	50.6	3335	2 H81702	adherence factor T
22	39	49.4	185	2 C90063	hypothetical prote
23	39	49.4	225	2 P97349	response regulator
24	39	49.4	233	1 C70106	conserved hypothet
25	39	49.4	257	2 F75084	hypothetical prote
26	39	49.4	362	2 I37459	Keratin Ha3-II, ty
27	39	49.4	404	2 JS0073	Keratin, 47.6K typ
28	39	49.4	412	1 KRSHL1	keratin, 48K type
29	39	49.4	416	2 S60034	keratin Ha1, type

30	39	49.4	786	2 E87565	conserved hypothet
31	39	49.4	828	2 D81522	DNA mismatch repai
32	39	49.4	828	2 C86608	DNA mismatch repai
33	39	49.4	828	2 G72017	DNA mismatch repai
34	39	49.4	1892	2 T18314	hypothetical prote
35	38	48.1	105	2 G71645	hypothetical prote
36	38	48.1	185	2 H96988	transcription regu
37	38	48.1	322	2 G64694	arginase - Helicob
38	38	48.1	324	2 B84150	hypothetical prote
39	38	48.1	362	2 S45094	cinnamyl-alcohol d
40	38	48.1	411	2 S45318	keratin 12 - rabbi
41	38	48.1	422	2 S59319	probable membrane
42	38	48.1	430	2 S05481	keratin 18, type I
43	38	48.1	476	2 S64425	hypothetical prote
44	38	48.1	488	2 JQ2257	nuclear antigen 21
45	38	48.1	488	2 T02207	protein 21D7 - com

ALIGNMENTS

RESULT 1

T22322  
hypothetical protein F46G10.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T22322  
R:Colles, L.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: Z19548  
A:Accession: T22322  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-336 <WIL>  
A:Cross-references: UNIPROT:O20478; UNIPARC:UPI000007A8AC; EMBL:Z50177; PIDN:CAA90544.1;  
A:Experimental source: clone F46G10  
C:Genetics:  
A:Gene: CBSP:F46G10.4  
A:Map position: X  
A:Introns: 16/3; 50/1; 80/3; 129/1; 162/2; 195/3; 216/3; 257/2; 290/3

Query Match 58.2%; Score 46; DB 2; Length 336;  
Best Local Similarity 60.0%; Pred. No. 4.8;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 EIEQYNLLLASVLD 15  
|||:|||||:  
Db 34 EAEQFLNLAAYAD 48

RESULT 2

T33888  
hypothetical protein Y39F10A.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T33888  
R:Leonard, S.; Graves, T.; Wilson, C.  
submitted to the EMBL Data Library, February 1999  
A:Description: The sequence of C. elegans cosmid Y39F10A.  
A:Reference number: Z21432  
A:Accession: T33888  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-309 <LEO>  
A:Cross-references: UNIPROT:Q9TYN8; UNIPARC:UPI000007DF80; EMBL:AFI25450; PIDN:AAD12817.  
A:Experimental source: strain Bristol N2; clone Y39F10A  
C:Genetics:  
A:Gene: CBSP:Y39F10A.3  
A:Map position: 2  
A:Introns: 33/1; 73/3; 139/2; 170/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein K06H6.2

Query Match 54.4%; Score 43; DB 2; Length 309;

Best Local Similarity 53.3%; Pred. No. 14;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 IEQYLNLLASYLDF 16  
Db 293 VEKYGGLKVANYLDF 307

RESULT 3  
T24114  
hypothetical protein R10D12.2 - Caenorhabditis elegans.  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T24114  
R:Percy, C.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19842  
A:Accession: T24114  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-408 <WTL>  
A:Cross-references: UNIPROT:Q9VKV3; UNIPARC:UPI000008028D; EMBL:Z81109; PIDN:CA803244.1;  
A:Experimental source: clone R10D12  
C:Genetics:  
A:Gene: CESP:R10D12.2  
A:Map position: 5  
A:Introns: 44/1; 101/1; 123/1; 151/1; 284/2; 351/3

Query Match 53.2%; Score 42; DB 2; Length 408;  
Best Local Similarity 53.8%; Pred. No. 28;  
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EIEQYLNLLASY 13  
Db 26 EIEKWNLILVSP 38

RESULT 4  
E86192  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: E86192  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E86192  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-471 <STO>  
A:Cross-references: UNIPROT:Q9WA46; UNIPARC:UPI000000A8FAC; GB:A8005172; NID:G6850308; PT  
C:Genetics:  
A:Map position: 1

Query Match 53.2%; Score 42; DB 2; Length 471;  
Best Local Similarity 62.5%; Pred. No. 32;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 EIEQYLNLLASYLDF 16  
Db 107 EITRYGNLLSASYKGF 122

RESULT 5  
564488

A:Accession: E97862  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-105 <KUR>  
A:Cross-references: UNIPROT:Q92G24; UNIPARC:UPI000000CC074; GB:AE006914; PIDN:AAL03839.1;  
C:Genetics:  
A:Gene: RC1301

Query Match 51.9%; Score 41; DB 2; Length 105;  
Best Local Similarity 58.3%; Pred. No. 10;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 EOYLNLLASYL 14  
Db 93 EDYLNLIINSYI 104  
|||||:|

RESULT 8  
S51444  
hypothetical protein YLR221c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein I8083.7  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
A:Accession: S51444  
R:Hallsorth, K.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmid 8083.  
A:Reference number: S51443  
A:Accession: S51444  
A:Molecule type: DNA  
A:Residues: 1-220 <HAL>  
A:Cross-references: UNIPROT:Q05942; UNIPARC:UPI0000052DF0; EMBL:U19027; NID:g609363; PIDN:  
C:Genetics:  
A:Gene: SGD:RSA3; MIPS:YLR221c  
A:Cross-references: SGD:S0004211  
A:Map position: 12R

Query Match 51.9%; Score 41; DB 2; Length 220;  
Best Local Similarity 61.5%; Pred. No. 22;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 EOYLNLLASYLD 15  
Db 162 EYLNLLFENYGD 174  
|||||:|

RESULT 9  
S37676  
glutamate dehydrogenase (EC 1.4.1.2) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: glutamic dehydrogenase; NAD-specific glutamate dehydrogenase; protein  
C:Species: Saccharomyces cerevisiae  
C:Date: 31-Dec-1993 #sequence\_revision 23-Mar-1995 #text\_change 05-Oct-2004  
A:Accession: S37676; S22282; S67774; S67778  
R:Boles, E.; Lehnert, W.; Zimmermann, F.K.  
Eur. J. Biochem. 217, 469-477, 1993  
A:Title: The role of the NAD-dependent glutamate dehydrogenase in restoring growth on glucose  
A:Reference number: S37676; MUID:94039072; PMID:7901008  
A:Accession: S37676  
A:Molecule type: DNA  
A:Residues: 1-1092 <BOL>  
A:Cross-references: UNIPROT:P33327; UNIPARC:UPI00001292E2; EMBL:X72015; NID:g396750; PIDN:  
R:Miller, S.M.; Magasanik, B.  
Mol. Cell. Biol. 11, 6229-6247, 1991  
A:Title: Role of the complex upstream region of the GDH2 gene in nitrogen regulation of  
A:Reference number: S22282; MUID:92049354; PMID:1682801  
A:Accession: S22282  
A:Molecule type: DNA  
A:Residues: 1-79 <ML>  
A:Cross-references: UNIPARC:UPI00001689AE; GB:S66436; NID:g238959; PIDN:AAB20327.1; PIDN:  
R:Schmidt, E.R.; Bahr, A.; Kraemer, C.; Hankeln, T.; Moeller-Rieker, S.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67756  
A:Accession: S67774

A:Molecule type: DNA  
A:Residues: 1-1092 <SCH>  
A:Cross-references: UNIPARC:UPI00001292E2; EMBL:Z74263; NID:g1431359; PIDN:CAA98793.1; PIDN:  
A:Experimental source: strain S288C  
R:Rasmussen, S.W.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67778  
A:Accession: S67778  
A:Molecule type: DNA  
A:Residues: 849-1092 <RAS>  
A:Cross-references: UNIPARC:UPI0000172054; EMBL:Z74263; GSPDB:GN00004; MIPS:YDL215C  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:GDH2; MIPS:YDL215C  
A:Cross-references: SGD:S0002374; MIPS:YDL215C  
A:Map position: 4L  
C:Superfamily: NAD-dependent glutamate dehydrogenase  
C:Keywords: NAD; oxidoreductase  
F:536,626/Binding site: substrate (Lys) #status predicted

Query Match 51.9%; Score 41; DB 1; Length 1092;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 EOYLNLLASYL 14  
Db 1049 ESYLVLLSSYL 1060  
|||||:|

RESULT 10  
F69785  
fructokinase homolog ydhr - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: F69785  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle,  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, I.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot,  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: F69785  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-299 <KUN>  
A:Cross-references: UNIPROT:Q05510; UNIPARC:UPI000005FFB8; GB:Z99107; GB:AL009126; NID:g  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ydhr  
C:Superfamily: fructokinase; glucose kinase homology  
F:50-186/Domain: glucose kinase homology <GKH>

Query Match 50.6%; Score 40; DB 2; Length 299;  
Best Local Similarity 53.3%; Pred. No. 44;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 IEQYVFKIMNSYLD 16  
Db 241 IYQYVFKIMNSYLD 255  
|||||:|

RESULT 11  
S31571

cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) - cottonwood (fragment)  
C:Species: Populus deltoides (cottonwood)  
C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 09-Jul-2004  
C:Accession: S31571  
R:van Doorselaere, J.; van Montagu, M.; Inze, D.  
submitted to the EMBL Data Library, January 1993  
A:Description: Isolation and characterization of poplar and alfalfa cinnamyl alcohol dehydrogenase homology  
A:Reference number: S31571  
A:Accession: S31571  
A:Molecule type: mRNA  
A:Residues: 1-341 <VAN>  
A:Cross-references: UNIPROT:P31657; UNIPARC:UPI0000174E44; EMBL:Z19568  
A:Experimental source: leaf  
C:Function:  
A:Pathway: lignin biosynthesis  
A:Note: last step in pathway  
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
C:Keywords: lignin biosynthesis; metalloprotein; oxidoreductase; zinc  
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>  
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 50.6%; Score 40; DB 2; Length 341;  
Best Local Similarity 60.0%; Pred. No. 50;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 EIEQYLNLLASYLD 15  
||||| : |||  
Db 109 EIEQYCNKKIWSYND 123

RESULT 12  
T09141  
cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) - cottonwood  
C:Species: Populus deltoides (cottonwood)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: T09141  
R:van Doorselaere, J.; Baucher, M.; Feuillet, C.; Boudet, A.M.; van Montagu, M.; Inze, Plant Physiol. Biochem. 33, 105-109, 1995  
A:Title: Isolation of cinnamyl alcohol dehydrogenase cDNAs from two important economic species  
A:Reference number: Z16583  
A:Accession: T09141  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-357 <VAN>  
A:Cross-references: UNIPROT:P31657; UNIPARC:UPI0000126DBA; EMBL:Z19568; NID:g288752; PID:100,103,106,114/Binding site: zinc, noncatalytic (Cys, His, Cys) #status predicted  
C:Function:  
A:Description: catalyzes the reversible oxidation of cinnamyl alcohol to cinnamaldehyde  
A:Pathway: lignin biosynthesis  
A:Note: last step in pathway  
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
C:Keywords: lignin biosynthesis; metalloprotein; oxidoreductase; zinc  
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>  
F:183-212/Region: beta-alpha-beta NADP nucleotide-binding fold  
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted  
F:100,103,106,114/Binding site: zinc, noncatalytic (Cys, His, Cys) #status predicted

Query Match 50.6%; Score 40; DB 2; Length 357;  
Best Local Similarity 60.0%; Pred. No. 53;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 EIEQYLNLLASYLD 15  
||||| : |||  
Db 109 EIEQYCNKKIWSYND 123

RESULT 13  
S31572  
cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) - alfalfa  
C:Species: Medicago sativa (alfalfa)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S31572  
R:van Doorselaere, J.; van Montagu, M.; Inze, D.

submitted to the EMBL Data Library, January 1993  
A:Description: Isolation and characterization of poplar and alfalfa cinnamyl alcohol dehydrogenase homology  
A:Reference number: S31571  
A:Accession: S31572  
A:Molecule type: mRNA  
A:Residues: 1-358 <VAN>  
A:Cross-references: UNIPROT:P31656; UNIPARC:UPI0000126DB7; EMBL:Z19573; NID:g19594; PID:NID:g19594; PID:N  
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc  
F:33-341/Domain: long-chain alcohol dehydrogenase homology <LADH>  
F:48,70,164/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 50.6%; Score 40; DB 2; Length 358;  
Best Local Similarity 60.0%; Pred. No. 53;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 EIEQYLNLLASYLD 15  
||||| : |||  
Db 110 EIEQYCNKKIWSYND 124

RESULT 14  
T01315  
hypothetical protein T14P8.16 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T01315  
R:Kalicki, J.; Elliott, G.; Cloud, J.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of A. thaliana T14P8.  
A:Reference number: Z14290  
A:Accession: T01315  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-771 <KAL>  
A:Cross-references: UNIPROT:O81298; UNIPARC:UPI000009E31F; EMBL:AF069298; NID:g3193282; I  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 328/3; 396/2  
A:Note: T14P8.16

Query Match 50.6%; Score 40; DB 2; Length 771;  
Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EIEQYLNLLASY 13  
||||| : |||  
Db 694 EIRQMNLSSH 706

RESULT 15  
ISECTB  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) chain B - Escherichia coli (strain K-12)  
N:Alternate names: DNA gyrase B chain; type II DNA topoisomerase B chain  
C:Species: Escherichia coli  
C:Date: 30-Sep-1988 #sequence\_revision 31-Oct-1997 #text\_change 05-Oct-2004  
C:Accession: D65172; A26444; A36953; C22168; A38344  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: D65172  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-804 <BLAT>  
A:Cross-references: UNIPROT:P06982; UNIPARC:UPI00001681E5; GB:AE000447; GB:U00096; NID:g  
A:Experimental source: strain K-12, substrain MG1655  
R:Adachi, T.; Mizuuchi, M.; Robinson, E.A.; Appella, E.; O'Des, M.H.; Gellert, M.; Mizuuc  
Nucleic Acids Res. 15, 771-784, 1987  
A:Title: DNA sequence of the E. coli gyrB gene: application of a new sequencing strategy.  
A:Reference number: A93674; MUID:87146392; PMID:3029692

A;Accession: A26444  
A;Molecule type: DNA  
A;Residues: 1-384, 'A', 386-804 <ADA>  
A;Cross-references: UNIPARC:UPI000016552C; GB:X04341; GB:X00870; NID:G41643; PIDN:CAA278  
R;Menzel, R.; Gellert, M.  
J. Bacteriol. 169, 1272-1278, 1987  
A;Title: Fusions of the Escherichia coli gyrA and gyrB control regions to the galactokinase gene.  
A;Reference number: A26953; MUID:87137287; PMID:3029031  
A;Accession: A26953  
A;Molecule type: DNA  
A;Residues: 1-23 <MEN>  
A;Cross-references: UNIPARC:UPI000016F1BF; GB:M15548; NID:G146307; PIDN:AAA23949.1; PID:  
R;Adachi, T.; Mizunuchi, K.; Menzel, R.; Gellert, M.  
Nucleic Acids Res. 12, 6389-6395, 1984  
A;Title: DNA sequence and transcription of the region upstream of the E. coli gyrB gene.  
A;Reference number: A22168; MUID:84237235; PMID:6089112  
A;Accession: C22168  
A;Molecule type: DNA  
A;Residues: 1-50, 'R', 52-55, 'L', 57-106 <AD2>  
A;Cross-references: UNIPARC:UPI00001730B3  
R;Tamura, J.K.; Gellert, M.  
J. Biol. Chem. 265, 21342-21349, 1990  
A;Title: Characterization of the ATP binding site on Escherichia coli DNA gyrase. Affinity  
A;Reference number: A38344; MUID:91065955; PMID:2174443  
A;Accession: A38344  
A;Molecule type: protein  
A;Residues: 93-126, 'X', 128-129 <TAM>  
A;Cross-references: UNIPARC:UPI00001730B4  
A;Note: Lys-103 and Lys-110 were shown to bind covalently to the ATP analog and DNA gyrase.  
C;Comment: DNA gyrase, which catalyzes DNA supercoiling and relaxing, is made up of two  
C;Genetics:  
A;Gene: gyrB  
A;Map position: 83 min  
C;Superfamily: Type II topoisomerase, subunit B  
C;Keywords: antibiotic resistance; ATP; DNA supercoiling; isomerase  
F;93-129/Region: ATP-binding

Query Match 50.6%; Score 40; DB 1; Length 804;  
Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 IEQYLNLLASYL 14  
Db 349 VEQMMNELLAEYL 361

Search completed: July 5, 2006, 19:25:15  
Job time: 16.6667 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:01:48 ; Search time 186 Seconds  
(without alignments)  
79.571 Million cell updates/sec

Title: US-10-774-928A-13  
Perfect score: 79  
Sequence: 1 EIEQYNLLLASVLD 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	60.8	458	Q7NB33 MYCGA	Q7nb33 mycoplasma
2	47	59.5	2982	Q4RXP0 TETNG	Q4rxp0 tetraodon n
3	46	58.2	336	Q20478 CAENORHABDI	Q20478 caenorhabdi
4	46	58.2	688	Q54MG6 DICTYOSTELI	Q54mg6 dictyosteli
5	45	57.0	266	Q4B3Q0 POLAROMONAS	Q4b3q0 polaromonas
6	45	57.0	988	Q813L6 PLAF7	Q813l6 plasmodium
7	44	55.7	245	Q4HIG6 CAMCO	Q4hig6 campylobact
8	44	55.7	519	Q5ZMD9 CHICK	Q5zmd9 gallus gall
9	44	55.7	619	Q7RD05 PLAYO	Q7rd05 plasmodium
10	44	55.7	1684	Q4UA79 THEAN	Q4ua79 theileria a
11	44	55.7	2543	Q8IBR8 PLAF7	Q8ibr8 plasmodium
12	43	54.4	248	Q5HUA7 CAMJR	Q5hua7 campylobact
13	43	54.4	309	Q9TVN8 CAEEL	Q9tyv8 caenorhabdi
14	43	54.4	375	Q65280 ASF	Q65280 african swi
15	43	54.4	493	Q6KITS PICTO	Q6kits picophilus
16	43	54.4	565	Q33JY3 METHU	Q33jy3 methanospir
17	43	54.4	641	Q2NIG3 9MOLU	Q2nig3 aster yello
18	43	54.4	1793	YCF1 LOTJA	Q9bbn6 lotus japon
19	42	53.2	253	Q2XC69 PSEPU	Q2xc69 pseudomonas
20	42	53.2	253	Q88F14 PSEPK	Q88f14 pseudomonas
21	42	53.2	256	Q52745 RUMPL	Q52745 ruminococcu
22	42	53.2	258	Q2NDQ2 9PROT	Q2ndq2 nitrosomona
23	42	53.2	267	Q7VJP4 HELHP	Q7vjp4 helicobacte
24	42	53.2	408	Q9XVK1 CAEEL	Q9xvk1 caenorhabdi
25	42	53.2	471	Q9MA46 ARATH	Q9ma46 arabidopsis
26	42	53.2	558	Q4P454 USTWA	Q4p454 ustilago ma
27	42	53.2	602	Q35P20 9BRAD	Q35p20 bradyrhizob
28	42	53.2	837	Q67R65 SYMTH	Q67r65 symbiococc
29	42	53.2	1613	1 RGYR METJA	Q58907 methanococc
30	42	53.2	1624	1 RGYR PYRHO	Q58530 pyrococcus
31	42	53.2	1711	1 RGYR_PYRKO	Q6f598 pyrococcus

32 42 53.2 1898 2 Q54KP2 DICDI  
33 42 53.2 3583 2 Q9V3N4 DROME  
34 42 53.2 3584 2 Q9U999 DROME  
35 41 51.9 105 1 Y1301 RICCN  
36 41 51.9 105 2 Q7PA41 RICSI  
37 41 51.9 105 2 Q68Y09 RICTY  
38 41 51.9 110 2 Q4UJV6 RICFE  
39 41 51.9 161 2 Q6MU33 MYCMS  
40 41 51.9 220 1 RSA3 YEAST  
41 41 51.9 233 2 Q36L25 MARRY  
42 41 51.9 270 2 Q2IED0 9DELT  
43 41 51.9 332 2 Q2US99 ASPOR  
44 41 51.9 332 2 Q61NZ5 CABER  
45 41 51.9 355 2 Q6RCS7 9ROSI

## ALIGNMENTS

## RESULT 1

Q7NB33 MYCGA PRELIMINARY; PRT; 458 AA.  
ID Q7NB33 MYCGA PRELIMINARY; PRT; 458 AA.  
AC Q7NB33;  
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.  
DT 15-DEC-2003, sequence version 1.  
DT 21-FEB-2006, entry version 15.  
DE Arca.  
GN Name=arca; OrderedLocusNames=MYCGA4470; ORFNames=MGA\_0105;  
OS Mycoplasma gallisepticum.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2096;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=R(low);  
RX MEDLINE=22830409; PubMed=12949158; DOI=10.1099/mic.0.26427-0;  
RA Papazisi L., Gorton T.S., Kutish G., Markham P.F., Browning G.F.,  
RA Nguyen D.K., Swartzell S., Madan A., Mahairas G., Geary S.J.;  
RT "The complete genome sequence of the avian pathogen Mycoplasma  
RT gallisepticum strain R(low).";  
RL Microbiology 149:2307-2316(2003).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AB015450; AAP56797.1; -; Genomic DNA.  
DR BioCyc; MGAL233150:MGA\_0105-MONOMER; -;  
DR GO; GO:0016990; P-arginine deiminase activity; IEA.  
DR GO; GO:0006527; P-arginine catabolism; IEA.  
DR InterPro; IPR003198; Amidino trans.  
DR InterPro; IPR003876; Arg.deiminase.  
DR Pfam; PF02274; Amidinotransf.; 1.  
DR PRINTS; PR01466; ARGDEIMINASE.  
KW Complete proteome.  
SQ SEQUENCE 458 AA; 52887 MW; 69CD34ED174AC2D CRC64;

Query Match 60.8%; Score 48; DB 2; Length 458;

Best Local Similarity 83.3%; Pred. No. 17;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 QYLNLLASVLD 15  
Db 112 QYLNLLASVLD 123

## \*RESULT 2

Q4RXP0 TETNG PRELIMINARY; PRT; 2982 AA.  
ID Q4RXP0 TETNG PRELIMINARY; PRT; 2982 AA.  
AC Q4RXP0;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome 11 SCAF14979, whole genome shotgun sequence.  
GN ORFNames=GSTENG00027341001;

OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed:15496914; DOI=10.1038/nature03025;  
RA Jallion O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bounau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthonard V., Jubin C., Caetelli V., Katinka M., Vacherie B.,  
RA Bionnet C., Skalli Z., Catolico L., Poulain J., De Berardinis V.,  
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Lardier G., Chaple C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.,  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
the early vertebrate proto-karyotype.";  
RL Nature 431:946-957(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DDJB databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DDJB whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CAA01014979; CAG06842.1; -; Genomic\_DNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0004930; F:G-protein coupled receptor activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0007156; F:homophilic cell adhesion; IEA.  
DR GO; GO:0007218; P:neuropeptide signaling pathway; IEA.  
DR InterPro; IPR000152; Asx hydroxyl\_S.  
DR InterPro; IPR002126; Cadherin.  
DR InterPro; IPR006210; EGF.  
DR InterPro; IPR000742; EGF\_3.  
DR InterPro; IPR001881; EGF\_Ca\_bd.  
DR InterPro; IPR002049; EGF\_laminin.  
DR InterPro; IPR013032; EGF\_like\_reg.  
DR InterPro; IPR000832; GPCR\_secretin.  
DR InterPro; IPR001879; hormone\_rcpt.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR000203; PKD\_cys\_rich.  
DR Pfam; PF00002; 7tm\_2; 1.  
DR Pfam; PF00028; Cadherin; 8.  
DR Pfam; PF00008; EGF; 5.  
DR Pfam; PF01825; GPS; 1.  
DR Pfam; PF02793; HRM; 1.  
DR Pfam; PF00053; Laminin\_EGF; 1.  
DR Pfam; PF02210; Laminin\_G\_2; 2.  
DR PRINTS; PR00205; CADHERIN.  
DR PRINTS; PR00249; GPCRSECRETIN.  
DR SMART; SM00112; CA; 8.  
DR SMART; SM00181; EGF; 5.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00180; EGF\_Lam; 1.  
DR SMART; SM00303; GPS; 1.  
DR SMART; SM00282; LamG; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00232; CADHERIN\_1; 6.  
DR PROSITE; PS02368; CADHERIN\_2; 8.  
DR PROSITE; PS00022; EGF\_1; 6.  
DR PROSITE; PS01186; EGF\_2; 3.

DR PROSITE; PS50026; EGF\_3; 5.  
DR PROSITE; PS01248; EGF\_LAM\_1; 1.  
DR PROSITE; PS50027; EGF\_LAM\_2; 1.  
DR PROSITE; PS50221; G\_PROTEIN\_RECEP\_F2\_3; 1.  
DR PROSITE; PS50261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
DR PROSITE; PS50221; GPS; 1.  
DR PROSITE; PS50025; LAM\_G\_DOMAIN; 2.  
SQ SEQUENCE 2982 AA; 328838 MW; F430A9574F9F82FF CRC64;  
  
Query Match 59.5%; Score 47; DB 2; Length 2982;  
Best Local Similarity 69.2%; Pred. No. 1.9e+02;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 3 EQVNLNLLASLYLD 15  
Db 1117 EQVNLNLLAQFLD 1129  
|||:||||:|  
|:|:|:|:|:|  
  
RESULT 3  
Q20478 CAEEL PRELIMINARY; PRT; 336 AA.  
AC Q20478;  
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.  
DT 01-NOV-1996, sequence version 1.  
DT 07-FEB-2006, entry version 30.  
DE Hypothetical protein.  
GN ORFNames=F46G10.4;  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;  
RG The C. elegans sequencing consortium;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; Z50177; CAA90544.1; -; Genomic\_DNA.  
DR PIR; T22322; T22322.  
DR Ensembl; F46G10.4; Caenorhabditis elegans.  
DR WormBase; WBGene0009798; F46G10.4.  
DR WormPep; F46G10.4; CE02240.  
DR GO; GO:0004806; F:triacylglycerol lipase activity; IEA.  
DR GO; GO:0006629; P:lipid metabolism; IEA.  
DR InterPro; IPR002921; Lipase\_3.  
DR Pfam; PF01764; Lipase\_3; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 336 AA; 37820 MW; AC665500AB670D89 CRC64;  
  
Query Match 58.2%; Score 46; DB 2; Length 336;  
Best Local Similarity 60.0%; Pred. No. 28;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 EIEQVNLNLLASLYLD 15  
Db 34 EAEQFLNLAAYAD 48  
|:|:|:|:|:|  
|:|:|:|:|:|  
  
RESULT 4  
Q54M06 DICDI PRELIMINARY; PRT; 688 AA.  
ID Q54M06\_DICDI  
AC Q54M06;  
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.  
DT 24-MAY-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Hypothetical protein (Fragment).  
GN ORFNames=DDB0218745;



```

OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Tuncgang B., Berriman M., Song J., Olsen R., Szatranski K., Xu Q.,
RA Suganb R., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Loulseghe H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Bartell B., Kuspa A.;
RA "The genome of the social amoeba Dictyostelium discoideum.";
RT Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AAF10100124; EAL64732.1; -; Genomic_DNA.
KW Hypothetical protein.
FT NON TER 688 688
SQ SEQUENCE 688 AA; 78241 MW; 33C58408004A0CC7 CRC64;

Query Match 58.2%; Score 46; DB 2; Length 688;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 EQYNLLLSYL 14
DB 146 ENYLLLVNVL 157

RESULT 5
Q4B300_9BURK PRELIMINARY; PRT; 266 AA.
AC Q4B300;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Twin-arginine translocating C-subunit.
GN Orlanames-BprodRPT_4563;
OS Polaronomas sp. JS666.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Polaronomas.
OX NCBI_TaxID=296591;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Polaronomas sp.
RT JS666.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
[2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-ORNL);

```

```

RA Larimer F., Land M.;
RT "Annotation of the draft assembly of Polaronomas sp. JS666.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AAF002000002; EAM40774.1; -; Genomic_DNA.
DR InterPro; IPR002033; Translocase.
DR Pfam; PF00902; TatC; 1.
DR PRINTS; PR01840; TATCFAMILY.
DR TIGRFAMs; TIGR00945; tatC; 1.
SQ SEQUENCE 266 AA; 28879 MW; F060941BBAA301D2 CRC64;

Query Match 57.0%; Score 45; DB 2; Length 266;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 EIEQVLLNLLLSYLD 16
DB 162 DIEEVNLFVGNFLAF 177

RESULT 6
Q8I3L6_PLAF7 PRELIMINARY; PRT; 988 AA.
AC Q8I3L6;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Hypothetical protein PFE1240w.
GN Name=PFE1240w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=3D7;
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C.M., Harris B., Harris D.,
RA Mungall K.L., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R.M., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
RN [2]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=3D7;
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC -----

```

```
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AL929354; CAD51612.1; -; Genomic DNA.
DR GO; GO:0003824; F: catalytic activity; IEA.
DR GO; GO:0005506; F: iron ion binding; IEA.
DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF04055; Radical_SAM; I.
KW Hypothetical protein.
SQ SEQUENCE 988 AA; 116754 MW; 356FB2B826AB8335 CRC64;
Query Match 57.0%; Score 45; DB 2; Length 988;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EIEQYNLLLASYLDF 16
Db 841 EIDSYNILQLGYPDF 856
RESULT 7
QHIG6_CAMCO PRELIMINARY; PRT; 245 AA.
AC Q4HIG6;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Hypothetical protein.
GN ORFNames=CC0927;
OS Campylobacter coli RM2228.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=306254;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RM2228;
RA Fouts D.E., Mongodin E.F., Mandrell R.E., Miller W.G., Rasko D.A.,
RA Jacques R.J., Brinkac L.M., DeBoy R.T., Parker C.T., Daugherty S.C.,
RA Dodson R.J., Durkin A.S., Madupu R.R., Sullivan S.A., Shetty J.U.,
RA Ayodeji M.A., Shvartsbeyn A.A., Schatz M.C., Badger J.H., Fraser C.M.,
RA Nelson K.E.;
RT "Major structural and novel potential virulence mechanisms from the
RT genomes of multiple Campylobacter species.";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AAFLO100001; EAL57653.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 245 AA; 29199 MW; E6387DA72E8F1F47 CRC64;
Query Match 55.7%; Score 44; DB 2; Length 245;
Best Local Similarity 47.4%; Pred. No. 45;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;
QY 2 IEQYNLL-----LASYLDF 16
Db 104 ISEYNILQLFLAGYIDF 122
RESULT 8
Q5ZMD9_CHICK PRELIMINARY; PRT; 519 AA.
ID Q5ZMD9_CHICK
AC Q5ZMD9;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 21-FEB-2006, entry version 9.
DE Hypothetical protein.
GN ORFNames=RCJMB04_2h4;
```

```
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE
RC STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayaishizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT gene function analysis.";
RL Genome Biol. 6:R6-R6(2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AJ719445; CAG31104.1; -; mRNA.
DR GO; GO:0005488; F: binding; IEA.
DR InterPro; IPR013143; PAM.
DR InterPro; IPR000717; PCI.
DR InterPro; IPR011990; TPR-like_helical.
DR Pfam; PF01399; PCI; 1.
DR SMART; SM00088; PINT; 1.
KW Hypothetical protein.
SQ SEQUENCE 519 AA; 59397 MW; 8D22F0307B685465 CRC64;
Query Match 55.7%; Score 44; DB 2; Length 519;
Best Local Similarity 64.3%; Pred. No. 1e+02;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 EIEQYNLLLASYL 14
Db 147 EVEAYLQLLEVIYL 160
RESULT 9
Q7RD05_PLAYO PRELIMINARY; PRT; 619 AA.
ID Q7RD05_PLAYO
AC Q7RD05;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Homo sapiens RIKEN cDNA 1600015H11 gene-related.
GN ORFNames=PY05621;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.B., Selegut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
```

```

DR EMBL; AABL01001803; EAA17664.1; -; Genomic_DNA.
DR HSP; O43172; IMZW.
DR InterPro; IPR003648; SFM.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINERPT.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD REPEATS 1; UNKNOWN_1.
DR PROSITE; PS0082; WD REPEATS 2; 3.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Complete proteome; Repeat; WD repeat.
SQ SEQUENCE 619 AA; 7467 MW; 7287DC126095BCEA CRC64;

Query Match 55.7%; Score 44; DB 2; Length 619;
Best Local Similarity 43.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 EIEQYLNLLASVLDLF 16
Db 104 EVBEFNILKKYIDF 119

RESULT 10
Q4UA79 THEAN PRELIMINARY; PRT; 1684 AA.
AC Q4UA79;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=TA07330;
OS Theileria annulata.
OS Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC Theileria.
OC NCBI_TaxID=5874;
RN [1];

NUCLEOTIDE SEQUENCE.
RC STRAIN=Ankara isolate clone C9;
RX PubMed=15994597; DOI=10.1258/jrm.98.7.320;
RA Pain A., Renauld H., Berriman M., Murphy L., Yeats C.A., Weir W.,
RA Kerhornou A., Aslett M., Bishop R., Bouchier C., Cochet M.,
RA Coulson R.M.R., Cronin A., de Villiers E.P., Fraser A., Fosker N.,
RA Gardner M., Goble A., Griffiths-Jones S., Harris D.E., Katzer F.,
RA Larke N., Lord A., Maser P., McKellar S., Mooney P., Morton F.,
RA Nene V., O'Neill S., Price C., Quail M.A., Rabinowitsch E.,
RA Rawlings N.D., Rutter S., Saunders D., Seeger K., Shah T., Squares R.,
RA Squares S., Tivey A., Walker A.R., Woodward J., Dobbelaere D.A.E.,
RA Langsley G., Rajandream M.A., McKeever D., Shields B., Tait A.,
RA Barrell B., Hall N.;
RT "Genome of the host-cell transforming parasite Theileria annulata
RT compared with T. parva.";
RL Science 309:131-133(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR940353; CAI76274.1; -; Genomic_DNA.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 1684 AA; 194143 MW; B15B653F0A7AFDDF CRC64;

Query Match 55.7%; Score 44; DB 2; Length 1684;
Best Local Similarity 75.0%; Pred. No. 3.5e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 EYLNLLASVYL 14
Db 534 EYLNLLQSYL 545

RESULT 11
Q8IBR8 PLAF7 PRELIMINARY; PRT; 2543 AA.
ID Q8IBR8 PLAF7
AC Q8IBR8;

01-MAR-2003, integrated into UniProtKB/TrEMBL.
01-MAR-2003, sequence version 1.
21-FEB-2006, entry version 18.
DE Cysteine repeat modular protein 2 homologue, putative.
GN Name=MAL7P1.92;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AL844506; CAD50930.1; -; Genomic_DNA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR006210; EGF.
DR InterPro; IPR013032; EGF_like_reg.
DR InterPro; IPR011641; GCC2_GCC3.
DR Pfam; PF07699; GCC2_GCC3; 3.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS01186; EGF; 2; 3.
SQ SEQUENCE 2543 AA; 296919 MW; 4353B37FF076D572 CRC64;

Query Match 55.7%; Score 44; DB 2; Length 2543;
Best Local Similarity 57.1%; Pred. No. 5.5e+02;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 EYLNLLASVLDLF 16
Db 657 DDYLELLSVYLDY 670

RESULT 12
Q5HUA7 CAMJR PRELIMINARY; PRT; 248 AA.
ID Q5HUA7;
AC Q5HUA7;
DT 15-FEB-2005, integrated into UniProtKB/TrEMBL.
DT 15-FEB-2005, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Hypothetical protein.
GN OrderedLocustNames=CJEL136;
OS Campylobacter jejuni (strain RM1221).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=195099;
RN [1];
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15660156; DOI=10.1371/journal.pbio.0030015;
RA Fouts D.E., Mongodin E.F., Mandrell R.E., Miller W.G., Rasko D.A.,
RA Ravel J., Brinkac L.M., DeBoy R.T., Parker C.T., Daugherty S.C.,
RA Dodson R.J., Durkin A.S., Madupu R., Sullivan S.A., Shetty J.U.,
RA Ayodeji M.A., Shvartsbeyn A., Schatz M.C., Badger J.H., Fraser C.M.,
RA Nelson K.E.;
RT "Major structural differences and novel potential virulence mechanisms
RT from the genomes of multiple Campylobacter species.";
RL PLoS Biol. 3:72-85(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000025; AAW35463.1; -; Genomic_DNA.
DR TIGR; CJEL136; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 248 AA; 29753 MW; D530FF310BC05233 CRC64;

Query Match 54.4%; Score 43; DB 2; Length 248;
Best Local Similarity 47.4%; Pred. No. 69;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

```

```
QY 2 IEQYLNLL----LASYLDF 16
Db 104 ISEYNVLQQLFLAGYIDF 122

RESULT 13
Q8YTN8_CABEL PRELIMINARY; PRT; 309 AA.
AC Q8YTN8;
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-MAY-2000, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE Hypothetical protein Y39F10A.3.
GN ORFName=Y39F10A.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998)
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF125450; AAD12817.1; -; Genomic_DNA.
DR PIR; T33888; T33888.
DR Ensembl; Y39F10A.3; Caenorhabditis elegans.
DR WormPep; Y39F10A.3; CE21727.
DR InterPro; IPR007744; DUF672.
DR Pfam; PF05050; DUF672; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 309 AA; 34939 MW; 0325BFDE32A09183 CRC64;

Query Match 54.4%; Score 43; DB 2; Length 309;
Best Local Similarity 53.3%; Pred. No. 87;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 IEQYLNLLLASYLDF 16
Db 293 VEKYGGLKVANYLDF 307

RESULT 14
Q65280_ASF PRELIMINARY; PRT; 375 AA.
AC Q65280;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 22.
DE T-lymphocyte surface antigen CD2 homologue.
GN Name=LMW8DR;
OS African swine fever virus (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10497;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Malawi L11-20/1;
RX MEDLINE=94167882; PubMed=7907198;
RA Borca M.V., Kutish G.F., Afonso C.L., Irueta P., Carrillo C., Brun A.,
RA Suesman M.D., Rock D.L.;
RT "An African swine fever virus gene with similarity to the T-lymphocyte
RT surface antigen CD2 mediates hemadsorption.";
RL Virology 199:463-468(1994).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----
CC EMBL; L15337; AAA42720.1; -; Genomic_DNA.
SQ SEQUENCE 375 AA; 42430 MW; 61A38A8DABDA132F CRC64;

Query Match 54.4%; Score 43; DB 2; Length 375;
Best Local Similarity 72.7%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 LNLILLASYLDF 16
Db 191 INTLLSSYLDF 201

RESULT 15
Q8KYT5_PICTO PRELIMINARY; PRT; 493 AA.
AC Q6KYT5;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Phytoene dehydrogenase (EC 1.3.-.-).
GN OrderedLocustNames=PTO1532;
OS Picrophilus torridus.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Picrophilaceae; Picrophilus.
OX NCBI_TaxID=82076;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=DSM 9790 / ATCC 700027;
RX PubMed=15184574; DOI=10.1073/pnas.0401356101;
RA Fuetterer O., Angelov A., Liesegang H., Gottschalk G., Schleper C.,
RA Schepers B., Dock C., Antranikian G., Liebl W.;
RT "Genome sequence of Picrophilus torridus and its implications for life
RT around pH 0.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9091-9096(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AE017261; AAT44117.1; -; Genomic_DNA.
DR GO; GO:0050660; FAD binding; IEA.
DR GO; GO:0016705; P:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0016117; P:carotenoid biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006076; FAD oxred.
DR InterPro; IPR008151; Phytoene_DH_rel.
DR Pfam; PF01593; Amino oxidase; 1.
DR Pfam; PF01266; DAO; 1.
DR ProDom; PD139017; Phytol dehydro; 1.
KW Complete proteome; Oxidoreductase.
SQ SEQUENCE 493 AA; 56625 MW; 4136BBF4D3588E16 CRC64;

Query Match 54.4%; Score 43; DB 2; Length 493;
Best Local Similarity 43.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 EIEQYLNLLASYLDF 16
Db 393 KLEELYLNENIYIDF 408

Search completed: July 5, 2006, 19:18:20
Job time : 189 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:44:12 ; Search time 147.667 Seconds  
(without alignments)  
49.540 Million cell updates/sec

Title: US-10-774-928a-13

Perfect score: 79  
Sequence: 1 ETEQYLNLLASVLD 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 8:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*  
9: Geneseqp2005s:\*  
10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	79	100.0	16	AAO26393	Aao26393 Psoriasis
2	79	100.0	16	ADVO4418	Adv04418 Leishmania
3	44	55.7	385	ABG27716	Abg27716 Novel hum
4	44	55.7	417	AAE333677	Aae333677 Human str
5	43	54.4	127	AAE72685	Aag72685 Murine OR
6	42	53.2	471	ADN72161	Adn72161 Thale cre
7	42	53.2	496	ABG99094	Abg99094 Methanoco
8	42	53.2	496	ADL23828	Adl23828 Methanoco
9	42	53.2	496	AAEF76704	Aef76704 Methanoco
10	42	53.2	1711	ADNA46550	Adn46550 Thermococ
11	42	53.2	3583	ABE64814	Abb64814 Drosophil
12	41	51.9	170	AAE06786	Aae06786 Human dua
13	41	51.9	220	ABR52601	Abr52601 Protein s
14	41	51.9	220	ADK63900	Adk63900 Disease t
15	41	51.9	241	AAE06776	Aae06776 Human dua
16	41	51.9	498	AAE04835	Aae04835 Human-SGP
17	41	51.9	509	AAE06775	Aae06775 Human dua
18	41	51.9	509	AAE07045	Aae07045 Human dua
19	41	51.9	509	AAE07044	Aae07044 Human dua
20	41	51.9	509	ADN76018	Adn76018 Human DSP
21	41	51.9	509	ADN75960	Adn75960 Human sig
22	41	51.9	509	ADY37105	Ady37105 Protein t
23	41	51.9	590	ABG34140	Abg34140 Antibody

24	41	51.9	590	7	ADC66319	Adc66319 Chinese h
25	41	51.9	590	7	ADL18099	Adl18099 Anti-lect
26	41	51.9	590	7	ADL35263	Adl35263 Chinese h
27	41	51.9	597	6	ADA33416	Ada33416 Acinetoba
28	40	50.6	128	3	AAG11666	Aag11666 Arabidops
29	40	50.6	128	3	AAG53339	Aag53339 Arabidops
30	40	50.6	128	8	ADT55865	Adt55865 Plant pol
31	40	50.6	128	10	AEF29162	Aef29162 Lead_Cere
32	40	50.6	128	10	AEF29160	Aef29160 Lead_Cere
33	40	50.6	180	9	AEC77038	Aec77038 Protein e
34	40	50.6	180	9	AEC77041	Aec77041 Protein e
35	40	50.6	239	6	ABM68652	Abm68652 Photorhab
36	40	50.6	301	6	ABM68681	Abm68681 Photorhab
37	40	50.6	336	6	ABM68968	Abm68968 Photorhab
38	40	50.6	500	6	ABU39986	Abu39986 Protein e
39	40	50.6	594	2	AAW36811	Aaw36811 Human GDP
40	40	50.6	594	8	ADO57047	Ado57047 Human fuc
41	40	50.6	594	9	ADX80743	Adx80743 Human fuc
42	40	50.6	748	5	ABP73773	Abp73773 Candida a
43	40	50.6	804	4	AAU34800	Aau34800 E. coli c
44	40	50.6	804	6	ABU28825	Abu28825 Protein e
45	40	50.6	966	4	ABB61168	Abb61168 Drosophil

## ALIGNMENTS

RESULT 1  
AAO26393  
ID AAO26393 standard; peptide; 16 AA.  
XX  
AC AAO26393;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 13.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
PN WO200274239-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
(AKIV-) AKIVA LLC.  
XX  
O'daly JA;  
XX  
WPI; 2003-018763/01.  
XX  
New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
Claim 1; Page 43; 56pp; English.  
XX  
The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention

```
SQ Sequence 16 AA;
Query Match 100.0%; Score 79; DB 6; Length 16;
Best Local Similarity 100.0%; Pred. No. 5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EIEQYLNLLASYLDF 16
   |||||
DB 1 EIEQYLNLLASYLDF 16
   |||||

RESULT 2
ADV04418
ID ADV04418 standard; peptide; 16 AA.
XX
AC ADV04418;
XX
DT 24-FEB-2005 (first entry)
XX
DE Leishmania immunotherapeutic peptide SEQ ID NO:13.
XX
KW T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;
KW psoriasis.
XX
OS Leishmania sp.
XX
PN US2004241168-A1.
XX
PD 02-DEC-2004.
XX
PF 09-FEB-2004; 2004US-00774928.
XX
PR 16-MAR-2001; 2001US-00809003.
PR 17-OCT-2003; 2003US-00687892.
XX
PA (ODAL/) ODALY J A.
PI Odaly JA;
XX
WPI; 2005-011563/01.
XX
Inhibiting selectively T-cell rolling in human, by administering compound
PT interfering with cutaneous lymphocyte antigen-E selectin interaction and
PT leukocyte function associated antigen-1 interaction.
XX
Claim 18; SEQ ID NO 13; 21pp; English.
XX
The invention relates to a novel method for selectively inhibiting T-cell
CC rolling in a human host, comprising administering a compound that
CC selectively interferes with the cutaneous lymphocyte antigen (CLA)-E
CC selectin interaction and leukocyte function associated antigen (LFA)-
CC 1/intercellular adhesion molecule (ICAM) and very late antigen
CC (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound
CC includes an immunotherapeutic agent, which comprises a purified protein
CC extract that is isolated by diethylaminoethyl Sephadex chromatography of
CC a Nonidet P-40 insoluble particulate antigen fraction derived from
CC isolated killed cells of anastigotes from one or more species of the
CC Leishmania genus, where the particulate antigen fraction is solubilized
CC with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied
CC to diethylaminoethyl Sephadex and is eluted with a solution comprising
CC 0.1 M sodium chloride, 8 M urea and 0.025 M
CC Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract
CC includes polypeptides having apparent molecular weights of 73, 80 and 82
CC kDa, after total reduction and alkylation. The species is L. amazonensis ,
CC L. venezuelensis , L. brasiliensis and/or L. chagasi. A compound of the
CC invention has antipsoriatic and immunostimulant activity, and inhibits T-
CC cell rolling by interfering with CLA-E selectin interaction and LFA-
CC 1/ICAM and VLA/VCAM interactions. The method is useful for selectively
CC inhibiting T-cell rolling in a human host, and for treating psoriasis.
CC The present sequence represents a Leishmania peptide used in the
CC invention.
XX
Sequence 16 AA;
Query Match 100.0%; Score 79; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EIEQYLNLLASYLDF 16
   |||||
DB 1 EIEQYLNLLASYLDF 16
   |||||

RESULT 3
ABG27716
ID ABG27716 standard; protein; 385 AA.
XX
AC ABG27716;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27707.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
WPI; 2001-639362/73.
DR N-PSDB; AAS91903.
XX
New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
Claim 20; SEQ ID NO 58075; 103pp; English.
XX
The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
Sequence 385 AA;
Query Match 55.7%; Score 44; DB 4; Length 385;
```

Best Local Similarity 73.3%; Pred. No. 1.1e+02; Mismatches 4; Indels 0; Gaps 0;  
Matches 11; Conservative 0;

QY 1 EIEQYLNLLASLYLD 15  
| | | | | | | | | |  
Db 172 ETMQSLNLLASLYLD 186

## RESULT 4

AAE33677  
ID AAE33677 standard; protein; 417 AA.

XX AC AAE33677;

XX DT 16-APR-2003 (first entry)

XX DE Human structural and cytoskeleton-associated protein (SCAP) #11.

XX Human; structural and cytoskeleton-associated protein; SCAP; leukaemia;  
KW cell proliferative disorder; actinic keratosis; cancer; gastroenteritis;  
KW cirrhosis; psoriasis; adenocarcinoma; Creutzfeldt-Jakob disease; stroke;  
KW sarcoma; neurological disorder; epilepsy; seasonal affective disorder;  
KW Huntington's disease; Alzheimer's disease; lymphoma; melanoma; myeloma;  
KW anxiety; schizophrenia; amnesia; viral infection; pneumonia; influenza;  
KW arteriosclerosis; smallpox; Colorado tick fever; rabies; gene therapy;  
KW neuroprotective; nootropic; neuroleptic; cytostatic; virucide.

XX OS Homo sapiens.

XX PN WO2002101009-A2.

XX PD 19-DEC-2002.

XX PF 06-JUN-2002; 2002WO-US017956.

XX PR 07-JUN-2001; 2001US-0296865P.

XX PR 08-JUN-2001; 2001US-0296878P.

XX PR 15-JUN-2001; 2001US-0298664P.

XX PR 21-JUN-2001; 2001US-0300149P.

XX PR 29-JUN-2001; 2001US-0302340P.

XX PR 06-JUL-2001; 2001US-0303481P.

XX PR 12-JUL-2001; 2001US-0305059P.

XX PR 21-DEC-2001; 2001US-0343557P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Tang TY, Warren BA, Honchell CD, Richardson TW, Elliott VS;

XX PI Wallia NK, Yue H, Batra S, Griffin JA, Baughn MR, Forsythe ID;

XX PI Burford N, Emerling BM, Sanjanwala MM, Khan FA, Lu DAM, Hafalia AJA;

XX PI Nguyen DB, Yang J, Li JX, Becha SD, Yao MG, Gietzen KJ, Luo W;

XX PI Lee EA, Ison CH, Lasek AKW;

XX WPI; 2003-148791/14.

XX DR N-PSDB; AAD51574.

XX PT New human structural and cytoskeleton-associated proteins and genes,

XX PT useful for diagnosing or treating cancers (e.g. leukemia or lymphoma),

XX PT viral infections (e.g. influenza) or neurological disorders (e.g.

XX PT epilepsy or stroke).

XX PS Claim 1; Col 181-182; 121pp; English.

XX The invention relates to human structural and cytoskeleton-associated

CC sarcoma), neurological disorders (e.g. epilepsy, seasonal affective  
CC disorder, Huntington's disease, stroke, Alzheimer's disease, anxiety,  
CC Creutzfeldt-Jakob disease, schizophrenia or amnesia) or viral infections  
CC (e.g. pneumonia, herpes, influenza, Colorado tick fever, smallpox, rabies  
CC or gastroenteritis). SCAP DNA is also used in gene therapy. The present  
CC sequence is human SCAP protein

XX Sequence 417 AA;

QY Query Match 55.7%; Score 44; DB 6; Length 417;

Db Best Local Similarity 73.3%; Pred. NO. 1.2e+02;

Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EIEQYLNLLASLYLD 15

Db 80 ETMQSLNLLASLYLD 94

## RESULT 5

AAE33677

ID AAG72685 standard; protein; 127 AA.

XX AC AAG72685;

XX DT 31-JUL-2001 (first entry)

XX DE Murine OR-like polypeptide query sequence, SEQ ID NO: 2367.

XX KW Mouse; olfactory receptor; OR; primary scent determination;

XX KW secondary scent determination; polypeptide library; odour receptor;

XX KW scent profile; scent fingerprint; scent representation.

XX OS Mus musculus.

XX PN WO200127158-A2.

XX PD 19-APR-2001.

XX PF 06-OCT-2000; 2000WO-US027582.

XX PR 08-OCT-1999; 99US-0158615P.

XX PR 24-FEB-2000; 2000US-0184809P.

XX PA (DIGI-) DIGISCENTS.

XX PA (YEDA) YEDA RES & DEV CO LTD.

XX PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

XX PT New polynucleotides which encode polypeptides involved in olfactory

XX PT sensation for identifying olfactory agonists and antagonists.

XX PS Example 6; Page 1603; 1857pp; English.

XX The present sequence is a polypeptide encoded by one of 105 newly mined

XX mouse genes. It was used as a query sequence in a database search of

XX olfactory receptor (OR)-like sequences. The invention relates to isolated

XX polynucleotides encoding polypeptides involved in olfactory sensation.

XX The polynucleotides can be used in screening for olfactory agonists and

XX antagonists. The methods allow for the determination of primary scents

XX and the identification of the odour receptors used to detect these

XX primary scents. The methods also enable determination of secondary scents

XX and the identification of combinations of odour receptors that are

XX involved in detecting such secondary scents. This enables the

XX construction of a scent representation (also called a scent fingerprint

XX or scent profile), which may be used to re-create and edit scents.

XX Libraries of olfactory receptors are useful for determining the

XX interaction pattern of a composition with the receptors, and can be used

XX for determining differences in the olfactory faculties of different

XX individuals

XX Sequence 127 AA;



Query Match 54.4%; Score 43; DB 4; Length 127;  
Best Local Similarity 66.7%; Pred. No. 46;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 QYLNLLASYLD 15  
| | | | : | | |  
DB 4 QYLNLLASYLD 15

RESULT 6  
ADN72161  
ID ADN72161 standard; protein; 471 AA.  
XX  
AC ADN72161;  
XX  
DT 15-JUL-2004 (first entry)  
XX  
DE Thale cress protein upregulated in E2Fa/Dpa expressing plants SeqID 56.  
XX  
KW plant; transgenic; E2Fa/Dpa transcription factor; growth regulator;  
KW animal feed product; thale cress; cell wall biosynthesis;  
KW nitrogen metabolism; carbon metabolism.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO2004035798-A2.  
XX  
PD 29-APR-2004.  
XX  
PF 20-OCT-2003; 2003WO-EP011659.  
XX  
PR 18-OCT-2002; 2002EP-00079408.  
XX  
PA (CROP-) CROPDESIGN NV.  
XX  
PI Inze D, De Veylder L, Vlieghe K;  
DR N-PSDB; ADN72160.  
XX  
PT Altering plant characteristics, useful for producing plants for enzyme or pharmaceutical production comprises modifying in a plant, expression of one or more nucleic acids and/or modifying level or activity of one or more proteins.  
PS Claim 1; SEQ ID NO 56; 134pp; English.  
XX  
CC This invention relates to a novel method for altering one or more plant characteristics. Specifically, it refers to identifying genes that are up - or down-regulated in transgenic plants overexpressing the heterodimeric E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to alter plant characteristics accordingly. The present invention describes generating transgenic plants for the production of growth regulators, enzymes, therapeutic, pharmaceuticals and animal feed products, where the altered plant characteristics are selected from increased yield or biomass, enhanced survival capacity, stress tolerance, plant architecture or physiology, altered endoreplication, biochemistry, signal transduction, storage lipid mobilisation and/or altered photosynthesis, each relative to the corresponding wild type plants. Accordingly, these sequences can also be useful as positive or negative selectable markers during transformation of cells or tissues. The identified genes play a role in a variety of biological processes such as DNA replication, cell wall biosynthesis, nitrogen and/or carbon metabolism or they function as transcription factors. This polypeptide sequence is thale cress protein expressed by a gene upregulated 1.3 fold or more in plants overexpressing the E2Fa/Dpa transcription factor, given in an exemplification of the invention.  
SQ Sequence 471 AA;

Query Match 53.2%; Score 42; DB 8; Length 471;  
Best Local Similarity 62.5%; Pred. No. 2.9e+02;

Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 EIEOYLNLLASYLDF 16  
| | | | : | | | |  
DB 107 EITRYGNLLSASYKGF 122

RESULT 7  
ABG99094  
ID ABG99094 standard; protein; 496 AA.  
XX  
AC ABG99094;  
XX  
DT 11-SEP-2003 (revised)  
DT 14-JAN-2003 (first entry)  
XX  
DE Methanococcus jannaschii intein Mja r-Gyr.  
XX  
KW intein; intein-catalysed cyclic peptide library; cancer; tumour cell; cardiovascular disease; obesity; neurological disorder; apoptosis; cell death; cell division; cell growth; arrhythmia potential;  
KW cardiomyocyte; heart failure; sarcolemmal calcium cycling; stroke; cell proliferation; atherosclerosis; metabolism; skin disorder;  
KW bone morphogenetic protein; endocrinology; infectious disease; viral infection; bacterial infection; diabetic ulcer; wound healing; keloid formation; skin connective tissue cell; antibiotic transport; drug resistance; cytostatic; anorectic; neuroprotective; cardiant; haemostatic.  
XX  
OS Methanocaldococcus jannaschii.  
XX  
PN WO200166565-A2.  
XX  
PD 13-SEP-2001.  
XX  
PF 06-MAR-2001; 2001WO-US007162.  
XX  
PR 06-MAR-2000; 2000US-0187130P.  
XX  
PA (RIGE-) RIGEL PHARM INC.  
XX  
PI Kinsella TM;  
XX  
DR WPI; 2001-589926/66.  
XX  
PT Fusion polypeptide for generating libraries of cyclic peptides in vivo, comprises terminal intein motifs and a random peptide.  
XX  
PS Disclosure; Fig 3T; 115pp; English.  
XX  
CC The present invention relates to methods and compositions utilising inteins to generate libraries of cyclic peptides in vivo. The compositions of the invention are useful for making cyclic peptides in vivo. Libraries of cells can be transformed with libraries of fusion for nucleic acids. Inteins-catalysed cyclic peptide libraries are useful for screening for an altered phenotype and identifying target molecules. The library is introduced into a cell, screened for an altered phenotype and target molecules that bind to the cyclic peptide are isolated. The method is useful for making cyclic peptides which retain biological activity, and for screening for cyclic peptides capable of altering the phenotype of a cell. The cyclic peptides are useful for altering cellular phenotypes and/or physiology, in screening assays to identify target molecules associated with changes in cellular phenotype or physiology and as drugs to treat a number of disease state, such as cancer, cardiovascular diseases, obesity and neurological disorders. The bioactive cyclic peptide is useful as the starting point for designing/synthesising derivative molecules with similar or more favourable properties for use as a drug and to pull out target molecules. By introducing random libraries into any tumour cell, peptides which induce apoptosis, cell death, loss of cell division or decreased cell growth can be identified and this method is useful in cancer applications. The methods are also useful in cardiovascular applications, to screen for diminished arrhythmia potential in cardiomyocytes, for



CC enhanced contractile properties of cardiomyocytes and diminish heart  
 CC failure potential and to identify agents that regulate the intracellular  
 CC and sarcolemmal calcium cycling in cardiomyocytes to prevent arrhythmias  
 CC and agents that diminish embolic phenomena in arteries and arterioles  
 CC leading to stroke. Candidate bioactive peptide libraries are inserted  
 CC into these cell types and their proliferation in response to specific  
 CC stimuli is monitored. Furthermore, the methods are also useful in  
 CC screening for decreases in atherosclerosis, in screens to regulate  
 CC obesity by controlling food intake mechanisms or reducing the responses  
 CC of the receptor signalling pathways that regulate metabolism, in  
 CC neurobiology applications, to screen for agonists of bone morphogenetic  
 CC proteins, in skin biology applications, endocrinology applications,  
 CC infectious disease applications including viral and bacterial infection,  
 CC to screen for cyclic peptides which block HIV-1 infection, regulation or  
 CC inhibition of keloid formation and wound healing for diabetic ulcers.  
 CC Candidate libraries are inserted into skin connective tissue cells and  
 CC bioactive peptides which promote the growth of these cells are isolated.  
 CC Further applications include screening for bioactive peptides that block  
 CC antibiotic transport mechanisms, drug toxicities and drug resistance,  
 CC improving the performance of existing or developmental drugs, and in  
 CC biotechnology. ABG99075-ABG99098 represent inteins from different  
 CC organisms. (Updated on 11-SEP-2003 to standardise OS field)  
 XX  
 XX Sequence 496 AA;

Query Match 53.2%; Score 42; DB 4; Length 496;  
 Best Local Similarity 53.8%; Pred. No. 3e+02;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EOYNLLASYLD 15  
 DB 280 ESYNALIAGYFD 292

RESULT 8  
 ADL23828  
 ID ADL23828 standard; protein; 496 AA.

XX AC ADL23828;  
 XX  
 XX 03-JUN-2004 (first entry)  
 XX  
 DE Methanococcus jannaschii intein Mja r-Gyr amino acid sequence.  
 XX  
 KW cyclic peptide; interacting protein inhibition; fusion nucleic acid;  
 KW intein motif; intein-catalysed cyclic peptide library;  
 KW cancer cell line; cardiovascular; neurobiology; bone biology;  
 KW biotechnology; skin biology; endocrinology; drug toxicity;  
 KW drug resistance; drug development.

XX OS Methanocaldococcus jannaschii.  
 XX WO2004019890-A2.  
 XX  
 XX PD 11-MAR-2004.

XX PF 30-AUG-2003; 2003WO-US027371.  
 XX  
 XX PR 30-AUG-2002; 2002US-00232758.  
 XX PR 23-APR-2003; 2003US-00422536.

XX (RIGE-) RIGEL PHARM INC.

XX PA Lorens JB, Pray T, Kinsella TM, Bennett MK;

XX WPI; 2004-239115/22.

XX Generating intracellular cyclic peptide libraries, useful in identifying  
 PT interaction of interacting proteins by providing cells comprising first  
 PT and second interacting protein, and contacting cells with library of  
 PT fusion nucleic acids.

XX PS Disclosure; Fig 3T; 156pp; English.

XX The present invention describes a method (M1) for identifying a cell  
 CC comprising cyclic peptides which prevent interaction of interacting  
 CC proteins. M1 comprises providing cells with first and second interacting  
 CC proteins and contacting cells with a library of fusion nucleic acids  
 CC comprising (from 5' to 3'): a nucleic acid encoding a C-terminal intein  
 CC motif, a peptide, and an N-terminal intein motif, under conditions such  
 CC that cyclic peptide is formed. Also described is a method (M2) for  
 CC isolating the nucleic acid that encodes the cyclic peptide which when  
 CC introduced into cell caused the change in ubiquitination of the molecule,  
 CC comprising contacting a cell culture with an intein-catalysed cyclic  
 CC peptide library such that cells in the culture receive not more than one  
 CC member of the library, and monitoring the cells for a change in  
 CC ubiquitination of a molecule. M1 is useful for identifying a cell  
 CC comprising a cyclic peptide which is capable of preventing interaction of  
 CC interacting proteins. M1 is useful in creating cell lines from cancer  
 CC patients and in cardiovascular, neurobiology, drug toxicities and drug  
 CC resistance applications. M1 is useful for identifying agents that  
 CC regulates intracellular and sarcolemmal calcium cycling in  
 CC cardiomyocytes, agents that diminish embolic phenomena in arteries and  
 CC arterioles, and for screening peptides that block transport mechanisms.  
 CC M1 is useful for screening enhanced contractile properties of  
 CC cardiomyocytes. M1 enables screening of a large number of cell types  
 CC under a wide variety of conditions. M1 enables improvement in the  
 CC performance of existing or developmental drugs. The present sequence is  
 CC used in the exemplification of the present invention.

XX Sequence 496 AA;

Query Match 53.2%; Score 42; DB 8; Length 496;  
 Best Local Similarity 53.8%; Pred. No. 3e+02;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EOYNLLASYLD 15  
 DB 280 ESYNALIAGYFD 292

RESULT 9  
 AEF76704  
 ID AEF76704 standard; protein; 496 AA.

XX AC AEF76704;

XX DT 06-APR-2006 (first entry)

XX DE Methanococcus jannaschii intein Mja r-Gyr amino acid sequence.

XX KW screening; protein interaction; protein production; intein; cytostatic;  
 KW cardiovascular-gen.; neuroprotective; cancer; cardiovascular disease;  
 KW neurodegenerative disease.

XX OS Methanocaldococcus jannaschii.

XX PN US2005260626-A1.

XX PD 24-NOV-2005.

XX PF 28-FEB-2005; 2005US-00069642.

XX PR 06-MAR-2000; 2000US-0187130P.

XX PR 06-MAR-2001; 2001US-00800770.

XX PR 30-AUG-2003; 2003WO-US027371.

XX (RIGE-) RIGEL PHARM INC.

XX PI Lorens JB, Pray TR, Kinsella TM, Bennett MK;

XX WPI; 2006-190833/20.

XX Identifying a cyclic peptide capable of modulating the ubiquitination of  
 PT a molecule, comprises contacting a cell culture with a nucleic acid

PT library encoding an intein-catalyzed cyclic peptide library.  
XX Disclosure; SEQ ID NO 20; 205pp; English.  
PS  
XX  
CC The invention relates to a screening method comprising contacting a  
CC culture of cells with a nucleic acid library encoding an intein-catalyzed  
CC cyclic peptide library, monitoring the cells for a change in  
CC ubiquitination of a molecule within a cell, and isolating the nucleic  
CC acid that encodes the cyclic peptide that when introduced into the cell  
CC caused the change in ubiquitination of the molecule. The method is useful  
CC for identifying a cyclic peptide capable of modulating ubiquitination of  
CC a molecule. The screening method (not necessarily limited to identifying  
CC only cyclic peptides that affect ubiquitination) can be used for  
CC identifying cyclic peptides for treating a wide variety of conditions  
CC e.g. cancer, cardiovascular disease, or neurodegenerative disease. The  
CC present sequence represents an intein amino acid sequence, which is used  
CC in the exemplification of the present invention.  
XX  
XX Sequence 496 AA;  
SQ

Query Match 53.2%; Score 42; DB 10; Length 496;  
Best Local Similarity 53.8%; Pred. No. 3e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 3 EOYLNLLASYLD 15  
|:|:|:|:|  
Db 280 ESYINALIAGYFD 292

RESULT 10  
ADN46550  
ID ADN46550 standard; protein; 1711 AA.  
XX  
AC ADN46550;  
XX  
XX 01-JUL-2004 (first entry)  
XX  
XX Thermococcus kodakaraensis KOD1 protein sequence SeqID428.  
XX  
XX gene disruption; gene targeting; marker gene; transformation;  
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;  
KW gene structure; gene function; enzyme activity; medicine;  
KW forensic science; food; drug inspection; molecular biology; immunology.  
XX  
XX Thermococcus kodakaraensis.  
OS  
XX  
XX WO2004022736-A1.  
FN  
XX  
XX 18-MAR-2004.  
PD  
XX  
XX 29-AUG-2003; 2003WO-IB003597.  
PF  
XX  
XX 30-AUG-2002; 2002JP-00319011.  
PR  
XX  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
PA  
XX  
XX Imanaka T, Atomi H;  
PI  
XX  
XX WPI; 2004-257583/24.  
DR  
XX  
XX Method for disrupting targeted gene in genome of organism particularly  
PT thermostable bacterium and with genome chips for analysis, applicable in  
PT studying gene structure and functions.  
PT  
XX  
XX Claim 9; SEQ ID NO 428; 59pp; Japanese.  
PS  
XX This invention relates to a novel method for targeting disruption of an  
CC arbitrary gene in a genome of an organism which comprises providing the  
CC whole sequential data of the genome of such organism, selecting at least  
CC 1 arbitrary region in the sequence, providing a vector that contains a  
CC sequence homologous with the selected region and a marker gene,  
CC transformation, and homologous recombination. The genome is preferably  
CC the genome of a hyperthermostable archaeobacterium, particularly

CC Thermococcus kodakaraensis KOD1. The method is for targeting the  
CC disruption of a gene in the genome of an organism, which is applicable in  
CC studying gene structure and functions as well as enzyme activities of  
CC encoded proteins and useful in medicine, forensic science, food or drug  
CC inspection, molecular biology and immunology. With this method, the  
CC identification of a gene at an arbitrary position in a genome can be achieved  
CC efficiently and reliably. The present sequence is that of a protein  
CC encoded by the genome of Thermococcus kodakaraensis which was derived  
CC using the method of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
SQ Sequence 1711 AA;  
Query Match 53.2%; Score 42; DB 8; Length 1711;  
Best Local Similarity 46.2%; Pred. No. 1.2e+03;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 3 EOYLNLLASYLD 15  
|:|:|:|:|  
Db 1235 BEYNAMIAGYFD 1247

RESULT 11  
ABB64814  
ID ABB64814 standard; protein; 3583 AA.  
XX  
AC ABB64814;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 21234.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
KW  
XX Drosophila melanogaster.  
OS  
XX WO200171042-A2.  
FN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-00614150.  
PR  
XX  
XX (PEKE) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
DR  
XX  
XX N-PSDB; ABL08917.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
PT  
XX  
XX Disclosure; SEQ ID NO 21234; 21pp + Sequence Listing; English.  
PS  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 3583 AA;  
SQ





GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:52:16 ; Search time 84.6667 Seconds  
(without alignments)  
87.537 Million cell updates/sec

Title: US-10-774-928A-13  
Perfect score: 79  
Sequence: 1 EIEQYLNLLASVLDLF 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:\*  
1: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
3: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
4: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /EMC\_Celerra\_SID33/ptodata/2/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	16	5	US-10-774-928-13
2	44	55.7	385	5	US-10-450-763-58075
3	44	55.7	417	5	US-10-479-764-11
4	42	53.2	496	3	US-09-800-770-20
5	42	53.2	496	4	US-10-422-536-20
6	42	53.2	496	6	US-11-011-880-20
7	42	53.2	496	6	US-11-069-642-20
8	42	53.2	3583	6	US-11-097-143-21234
9	41	51.9	57	4	US-10-437-963-178257
10	41	51.9	111	4	US-10-425-115-222175
11	41	51.9	170	3	US-09-775-925-32
12	41	51.9	170	5	US-10-962-126-32
13	41	51.9	220	4	US-10-477-369-87
14	41	51.9	241	3	US-09-775-925-8
15	41	51.9	241	5	US-10-962-126-8
16	41	51.9	254	4	US-10-425-115-266235
17	41	51.9	355	6	US-11-188-298-15448
18	41	51.9	357	6	US-11-087-099-3020
19	41	51.9	498	4	US-10-168-506-15
20	41	51.9	498	5	US-10-838-181-15
21	41	51.9	509	3	US-09-775-925-6
22	41	51.9	509	3	US-09-555-732-14
23	41	51.9	509	4	US-10-444-795B-785
24	41	51.9	509	5	US-10-962-126-6
25	41	51.9	509	3	US-09-971-773-73
26	41	51.9	590	4	US-10-327-663-63
27	41	51.9	590	4	US-10-409-616-90

28	41	51.9	590	5	US-10-409-611-20	Sequence 20, Appl
29	41	51.9	590	6	US-11-131-212-73	Sequence 73, Appl
30	41	51.9	590	6	US-11-218-473-73	Sequence 73, Appl
31	41	51.9	590	6	US-11-240-579-73	Sequence 73, Appl
32	41	51.9	590	6	US-11-287-324-73	Sequence 73, Appl
33	41	51.9	590	6	US-11-287-359-73	Sequence 73, Appl
34	41	51.9	784	6	US-11-188-298-8237	Sequence 8237, Ap
35	40	50.6	98	4	US-10-425-115-310606	Sequence 310606,
36	40	50.6	105	4	US-10-767-701-58372	Sequence 58372, A
37	40	50.6	108	4	US-10-424-599-249754	Sequence 249754,
38	40	50.6	128	5	US-10-739-930-5942	Sequence 5942, Ap
39	40	50.6	128	6	US-11-172-740-1360	Sequence 1360, Ap
40	40	50.6	128	6	US-11-172-740-1362	Sequence 1362, Ap
41	40	50.6	335	6	US-11-188-298-21671	Sequence 21671, A
42	40	50.6	341	6	US-11-188-298-20692	Sequence 20692, A
43	40	50.6	357	6	US-11-188-298-13772	Sequence 13772, A
44	40	50.6	357	6	US-11-188-298-14546	Sequence 14546, A
45	40	50.6	358	6	US-11-188-298-9968	Sequence 9968, Ap

## ALIGNMENTS

RESULT 1  
US-10-774-928-13  
; Sequence 13, Application US/10774928  
; Publication No. US20040341168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; FILE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-13

Query Match 100.0%; Score 79; DB 5; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIEQYLNLLASVLDLF 16  
DB 1 EIEQYLNLLASVLDLF 16

RESULT 2  
US-10-450-763-58075  
; Sequence 58075, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 58075  
; LENGTH: 385  
; TYPE: PRT  
; ORGANISM: Homo sapiens



1990

Best Local Similarity 53.8%; Pred. No. 55;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EIQYNLLLASYL 14  
Db 25 VEIYMNILHSYL 37

RESULT 10  
US-10-425-115-222175  
; Sequence 222175, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 222175  
; LENGTH: 111  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(111)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_134213C.1.pep  
US-10-425-115-222175

Query Match 51.9%; Score 41; DB 4; Length 170;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EOYNLLLASYLDF 16  
Db 17 QHYNLLLASVHSF 30

RESULT 11  
US-09-775-925-32  
; Sequence 32, Application US/09775925  
; Patent No. US20010049358A1  
; GENERAL INFORMATION:  
; APPLICANT: Luche, Ralf M.  
; APPLICANT: Wei, Bo  
; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY  
; FILE REFERENCE: 200125.420  
; CURRENT APPLICATION NUMBER: US/09/775,925  
; CURRENT FILING DATE: 2001-02-01  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-775-925-32

Query Match 51.9%; Score 41; DB 3; Length 170;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EIEQYNLLLASYL 14  
Db 157 QLEEQGILLASFL 170

RESULT 12  
US-10-962-126-32  
; Sequence 32, Application US/10962126  
; Publication No. US20050058650A1  
; GENERAL INFORMATION:  
; APPLICANT: Luche, Ralf M.  
; APPLICANT: Wei, Bo  
; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY  
; FILE REFERENCE: 200125.420C1  
; CURRENT APPLICATION NUMBER: US/10/962,126  
; CURRENT FILING DATE: 2004-10-08  
; PRIOR APPLICATION NUMBER: US 09/775,925  
; PRIOR FILING DATE: 2001-02-01  
; PRIOR APPLICATION NUMBER: US 60/179,886  
; PRIOR FILING DATE: 2000-02-02  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-962-126-32

Query Match 51.9%; Score 41; DB 5; Length 170;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EIEQYNLLLASYL 14  
Db 157 QLEEQGILLASFL 170

RESULT 13  
US-10-477-369-87  
; Sequence 87, Application US/10477369  
; Publication No. US20040167066A1  
; GENERAL INFORMATION:  
; APPLICANT: Marzloch et al.  
; TITLE OF INVENTION: CLEAVAGE AND POLYADENYLATION COMPLEX OF PRECURSOR MRNA  
; FILE REFERENCE: 224967  
; CURRENT APPLICATION NUMBER: US/10/477,369  
; CURRENT FILING DATE: 2003-11-12  
; PRIOR APPLICATION NUMBER: PCT/EP02/05359  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: EP 01111774.4  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: EP 01130253.6  
; PRIOR FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 87  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-477-369-87

Query Match 51.9%; Score 41; DB 4; Length 220;  
Best Local Similarity 61.5%; Pred. No. 2.2e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 EOYNLLLASYL 15  
Db 162 BEYNLLFENYGD 174

RESULT 14  
US-09-775-925-8  
; Sequence 8, Application US/09775925  
; Patent No. US20010049358A1  
; GENERAL INFORMATION:  
; APPLICANT: Luche, Ralf M.  
; APPLICANT: Wei, Bo



; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY  
; TITLE OF INVENTION: PHOSPHATASES  
; FILE REFERENCE: 200125.420  
; CURRENT APPLICATION NUMBER: US/09/775,925  
; CURRENT FILING DATE: 2001-02-01  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-775-925-8

Query Match 51.9%; Score 41; DB 3; Length 241;  
Best Local Similarity 50.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIEQYLNLLASFL 14  
:|:|:|:|:|:|  
Db 177 QLEEQGILLASFL 190

RESULT 15  
US-10-962-126-8  
; Sequence 8, Application US/10962126  
; Publication No. US20050058650A1  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Bo  
; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY  
; TITLE OF INVENTION: PHOSPHATASES  
; FILE REFERENCE: 200125.420C1  
; CURRENT APPLICATION NUMBER: US/10/962,126  
; CURRENT FILING DATE: 2004-10-08  
; PRIOR APPLICATION NUMBER: US 09/775,925  
; PRIOR FILING DATE: 2001-02-01  
; PRIOR APPLICATION NUMBER: US 60/179,886  
; PRIOR FILING DATE: 2000-02-02  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-962-126-8

Query Match 51.9%; Score 41; DB 5; Length 241;  
Best Local Similarity 50.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIEQYLNLLASFL 14  
:|:|:|:|:|:|  
Db 177 QLEEQGILLASFL 190

Search completed: July 5, 2006, 20:57:20  
Job time : 85.6667 secs

**THIS PAGE BLANK (USPTO)**



```
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42014
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-42014

Query Match          49.4%; Score 39; DB 6; Length 152;
Best Local Similarity 53.3%; Pred. No. 9.6;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 EIEQYNLLLSYLD 15
Db 97 EIEQYISSQLAEQVD 111

RESULT 3
US-10-471-571A-3746
; Sequence 3746, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWinn99, version 1.03
; SEQ ID NO 3746
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(185)
; OTHER INFORMATION: hypothetical protein
US-10-471-571A-3746

Query Match          49.4%; Score 39; DB 6; Length 185;
Best Local Similarity 64.3%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 IEQYNLLLSYLD 15
Db 166 IEAYRNLLNIYLD 179

RESULT 4
US-10-449-902-39990
; Sequence 39990, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205V1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39990
```

```
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-39990

Query Match          49.4%; Score 39; DB 6; Length 648;
Best Local Similarity 57.1%; Pred. No. 46;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EIEQYNLLLSYLD 14
Db 255 EIQYLLHLLVIYL 268

RESULT 5
US-10-953-349-8457
; Sequence 8457, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 8457
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-8457

Query Match          48.1%; Score 38; DB 6; Length 363;
Best Local Similarity 57.1%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EQYNLLLSYLD 16
Db 178 ETLNLLLRNYLHY 191

RESULT 6
US-10-449-902-54430
; Sequence 54430, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205V1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54430
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-54430

Query Match          48.1%; Score 38; DB 6; Length 372;
Best Local Similarity 57.1%; Pred. No. 37;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EQYNLLLSYLD 16
```

Db 201 ETLNLLRLNYLHY 214

## RESULT 7

US-10-953-349-8456  
; Sequence 8456, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 8456  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-8456

Query Match 48.1%; Score 38; DB 6; Length 381;  
Best Local Similarity 57.1%; Pred. No. 38;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EQYNLLLSYLDLF 16  
| | | | | : | | :  
Db 196 ETLNLLRLNYLHY 209

## RESULT 8

US-10-953-349-8455  
; Sequence 8455, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 8455  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-8455

Query Match 48.1%; Score 38; DB 6; Length 387;  
Best Local Similarity 57.1%; Pred. No. 39;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EQYNLLLSYLDLF 16  
| | | | | : | | :  
Db 202 ETLNLLRLNYLHY 215

## RESULT 9

US-11-105-233-181  
; Sequence 181, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Thiagalagam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; FILE REFERENCE: 1657/2001  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; CURRENT FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 181

; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-105-233-181

Query Match 48.1%; Score 38; DB 7; Length 430;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 EIEQYNLLLSYLD 15  
| | | | | : | | :  
Db 82 ETMQSLNDRSLSYLD 96

## RESULT 10

US-10-953-349-942  
; Sequence 942, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 942  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-942

Query Match 48.1%; Score 38; DB 6; Length 481;  
Best Local Similarity 57.1%; Pred. No. 49;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EOYNLLLSYLDLF 16  
| | | | | : | | :  
Db 195 ETLNLLRLNYLHY 208

## RESULT 11

US-10-449-902-32091  
; Sequence 32091, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agrobiological Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF  
; FILE REFERENCE: MOA-A0205Y1-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 56791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32091  
; LENGTH: 486  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-449-902-32091

Query Match 48.1%; Score 38; DB 6; Length 486;  
Best Local Similarity 57.1%; Pred. No. 50;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EOYNLLLSYLDLF 16  
| | | | | : | | :  
Db 195 ETLNLLRLNYLHY 208

```
Db      200 ETLNLLRLNYLHY 213

RESULT 12
US-10-953-349-941
; Sequence 941, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 941
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-941

Query Match      48.1%; Score 38; DB 6; Length 487;
Best Local Similarity 57.1%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 4; Indels 4; Gaps 0;

QY      3 EQYNLLLASLYLDF 16
      | | | | | : | | :
Db      201 ETLNLLRLNYLHY 214

RESULT 13
US-10-449-902-53883
; Sequence 53883, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53883
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53883

Query Match      48.1%; Score 38; DB 6; Length 487;
Best Local Similarity 57.1%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 4; Indels 4; Gaps 0;

QY      3 EQYNLLLASLYLDF 16
      | | | | | : | | :
Db      201 ETLNLLRLNYLHY 214

RESULT 14
US-10-953-349-940
; Sequence 940, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
```

```
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 940
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-940

Query Match      48.1%; Score 38; DB 6; Length 506;
Best Local Similarity 57.1%; Pred. No. 52;
Matches 8; Conservative 2; Mismatches 4; Indels 4; Gaps 0;

QY      3 EQYNLLLASLYLDF 16
      | | | | | : | | :
Db      220 ETLNLLRLNYLHY 233

RESULT 15
US-10-449-902-45252
; Sequence 45252, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45252
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-45252

Query Match      48.1%; Score 38; DB 6; Length 838;
Best Local Similarity 56.2%; Pred. No. 90;
Matches 9; Conservative 1; Mismatches 6; Indels 6; Gaps 0;

QY      1 BIEQYNLLLASLYLDF 16
      | | | | | : | | :
Db      710 EANQCLNVLAKYDFF 725

Search completed: July 5, 2006, 20:57:48
Job time : 8.33333 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:24:38 ; Search time 28.333 Seconds  
(without alignments)  
49.429 Million cell updates/sec

**Title:** US-10-774-928A-13  
**Perfect score:** 79  
**Sequence:** 1 EIEQYLNLLASYLDF 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Database :
Issued Patents AA..
1: /EMC-Celerra SIDS3/ptodata/2/iaa/5 COMB pep.*
2: /EMC-Celerra SIDS3/ptodata/2/iaa/6 COMB pep.*
3: /EMC-Celerra SIDS3/ptodata/2/iaa/7 COMB pep.*
4: /EMC-Celerra SIDS3/ptodata/2/iaa/H COMB pep.*
5: /EMC-Celerra SIDS3/ptodata/2/iaa/PCTUS COMB pep.*
6: /EMC-Celerra SIDS3/ptodata/2/iaa/RE COMB pep.*
7: /EMC-Celerra SIDS3/ptodata/2/iaa/backfiles1 pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	79	100.0	16	2	US-09-809-003A-13	Sequence 13, Appl
2	42	53.2	117	2	US-09-248-796A-22931	Sequence 22931, A
3	42	53.2	496	2	US-09-800-170-20	Sequence 20, Appl
4	41	51.9	170	2	US-09-775-925-32	Sequence 32, Appl
5	41	51.9	241	2	US-09-775-925-8	Sequence 8, Appl
6	41	51.9	509	2	US-09-955-732A-14	Sequence 14, Appl
7	41	51.9	509	2	US-09-775-925-6	Sequence 6, Appl
8	41	51.9	590	2	US-09-971-773-73	Sequence 73, Appl
9	41	51.9	597	2	US-09-328-352-4703	Sequence 4703, Ap
10	41	51.9	1092	2	US-09-538-092-122	Sequence 122, App
11	40	50.6	318	2	US-09-248-796A-18266	Sequence 18266, A
12	40	50.6	453	2	US-09-248-796A-14922	Sequence 14922, A
13	40	50.6	594	2	US-08-826-964-2	Sequence 2, Appl
14	40	50.6	606	2	US-09-949-016-7691	Sequence 7691, Ap
15	39.5	50.0	192	2	US-09-248-796A-28177	Sequence 28177, A
16	39	49.4	253	2	US-09-248-796A-21157	Sequence 21157, A
17	39	49.4	328	2	US-09-270-767-58019	Sequence 58019, A
18	39	49.4	339	2	US-09-198-452A-1013	Sequence 1013, Ap
19	39	49.4	376	2	US-09-949-016-8106	Sequence 8106, Ap
20	39	49.4	377	2	US-09-949-016-8141	Sequence 8141, Ap
21	39	49.4	389	2	US-09-248-796A-15248	Sequence 15248, A
22	39	49.4	404	2	US-09-949-016-6202	Sequence 6202, Ap
23	39	49.4	416	2	US-09-949-016-6201	Sequence 6201, Ap
24	39	49.4	429	2	US-09-248-796A-19384	Sequence 19384, A
25	39	49.4	433	2	US-09-949-016-10900	Sequence 10900, A
26	39	49.4	436	2	US-09-949-016-6983	Sequence 6983, Ap

## ALIGNMENTS

## RESULT 1

```

RES001.1
US-09-809-003A-13
; Sequence 13, Application US/09809003A
; Patent No. 6673351
; GENERAL INFORMATION:
; APPLICANT: O'Daly, Jose
; TITLE OF INVENTION: Compositions and
; TITLE OF INVENTION: Psoriasis
; FILE REFERENCE: 302178
; CURRENT APPLICATION NUMBER: US/09/80
; CURRENT FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Leishmania
US-09-809-003A-13

```

```
Query Match      100.0%; Score 79; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 5.9e-07;
Matches 16: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 EIEQYNLLASYLDF 16  
|||  
Db 1 EIEQYNLLASYLDF 16

## RESULT 2

```

US-09-248-796A-22931
; Sequence 22931, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO
; TITLE OF INVENTION: FOR DIAGNOSTICS AND T
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIORITY APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIORITY APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22931
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
;

```

```
; NAME/KEY: UNSURE
; LOCATION: (6), (66), (73), (75), (79)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-22931

Query Match      53.2%; Score 42; DB 2; Length 117;
Best Local Similarity 31.2%; Pred. No. 11;
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 BIEQYNLLLASYLDF 16
   ::||::||:|
Db 89 KLRNYIILSPYIEF 104

RESULT 3
US-09-800-170-20
; Sequence 20, Application US/09800170
; Patent No. 6481667
; GENERAL INFORMATION:
; APPLICANT: Kinsella, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 20
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-09-800-170-20

Query Match      53.2%; Score 42; DB 2; Length 496;
Best Local Similarity 53.8%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EOYNLLLASLYLD 15
   ||::||:|
Db 280 ESYNALIAGYDF 292

RESULT 4
US-09-775-925-32
; Sequence 32, Application US/09775925
; Patent No. 6902919
; GENERAL INFORMATION:
; APPLICANT: Lucche, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY
; FILE REFERENCE: 200125.420
; CURRENT APPLICATION NUMBER: US/09/775,925
; CURRENT FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-925-32

Query Match      51.9%; Score 41; DB 2; Length 170;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 BIEQYNLLLASYL 14
   ::||:|
Db 157 QLEEQGILLASFL 170

RESULT 5
US-09-775-925-6
; Sequence 6, Application US/09775925
; Patent No. 6902919
; GENERAL INFORMATION:
; APPLICANT: Lucche, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY
; FILE REFERENCE: 200125.420
; CURRENT APPLICATION NUMBER: US/09/775,925
; CURRENT FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-925-6

Query Match      51.9%; Score 41; DB 2; Length 509;
Best Local Similarity 50.0%; Pred. No. 86;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 BIEQYNLLLASYL 14
   ::||:|
Db 445 QLEEQGILLASFL 458

RESULT 6
US-09-775-925-8
; Sequence 8, Application US/09775925
; Patent No. 6902919
; GENERAL INFORMATION:
; APPLICANT: Lucche, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY
; FILE REFERENCE: 200125.420
; CURRENT APPLICATION NUMBER: US/09/775,925
; CURRENT FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-925-8

Query Match      51.9%; Score 41; DB 2; Length 241;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 BIEQYNLLLASYL 14
   ::||:|
Db 177 QLEEQGILLASFL 190

RESULT 7
US-09-955-732A-14
; Sequence 14, Application US/09955732A
; Patent No. 6825021
; GENERAL INFORMATION:
; APPLICANT: Lucche, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-15 DUAL-SPECIFICITY PHOSPHATASE
; FILE REFERENCE: 200125.433
; CURRENT APPLICATION NUMBER: US/09/955,732A
; CURRENT FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-732A-14

Query Match      51.9%; Score 41; DB 2; Length 509;
Best Local Similarity 50.0%; Pred. No. 86;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 BIEQYNLLLASYL 14
   ::||:|
Db 445 QLEEQGILLASFL 458

RESULT 7
US-09-775-925-6
; Sequence 6, Application US/09775925
; Patent No. 6902919
; GENERAL INFORMATION:
; APPLICANT: Lucche, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-12 AND DSP-13 DUAL-SPECIFICITY
; FILE REFERENCE: 200125.420
; CURRENT APPLICATION NUMBER: US/09/775,925
; CURRENT FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-925-6

Query Match      51.9%; Score 41; DB 2; Length 170;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 BIEQYNLLLASYL 14
   ::||:|
Db 157 QLEEQGILLASFL 170
```





```
; Sequence 14922, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14922
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14922

Query Match      50.6%; Score 40; DB 2; Length 453;
Best Local Similarity 58.3%; Pred. No. 1.1e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy      3 IEQYNLLASYL 14
Db      232 EQYLSLLISKFL 243

RESULT 13
US-08-826-964-2
; Sequence 2, Application US/08826964
; Patent No. 6033663
; GENERAL INFORMATION:
; APPLICANT: Ketcham, Catherine M.
; TITLE OF INVENTION: Nucleic Acids Encoding GDP-Fucose
; TITLE OF INVENTION: Pyrophosphorylase
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,964
; FILING DATE: 09-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/015,241
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 014137-009910US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 594 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-826-964-2

Query Match      50.6%; Score 40; DB 2; Length 594;
```

```
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      2 IEQYNLLASYLDF 16
Db      140 IQOMLEKLAMYIDF 154

RESULT 14
US-09-949-016-7691
; Sequence 7691, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7691
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7691

Query Match      50.6%; Score 40; DB 2; Length 606;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      2 IEQYNLLASYLDF 16
Db      152 IQOMLEKLAMYIDF 166

RESULT 15
US-09-248-796A-28177
; Sequence 28177, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 28177
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-28177

Query Match      50.0%; Score 39.5; DB 2; Length 192;
Best Local Similarity 64.3%; Pred. No. 51;
Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Qy      1 EIEQYNLLASYL 14
Db      68 EIEYVNLVI-SYL 80
```

Search completed: July 5, 2006, 19:26:44  
Job time : 29.3333 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:09:18 ; Search time 19.4175 Seconds  
(without alignments)  
79.283 Million cell updates/sec

Title: US-10-774-928a-14

Perfect score: 77

Sequence: 1 STMQELNSRLASYLDK 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*

1: Pirl.\*  
2: Pirl2.\*  
3: Pirl3.\*  
4: Pirl4.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	100.0	622	2 I37984	keratin 9, type I,
2	66	85.7	423	2 I59463	keratin, type I, c
3	65	84.4	432	2 S30433	keratin 17, type I
4	64	83.1	327	2 S04511	keratin 3, type I
5	64	83.1	400	1 KRH09	keratin 19, type I
6	64	83.1	401	2 S57657	keratin 19 - potot
7	64	83.1	403	2 JQ0028	cytokeratin 19 - m
8	64	83.1	411	2 S45318	keratin 12 - rabbi
9	64	83.1	437	2 A55682	keratin 13, type I
10	64	83.1	452	2 I49595	cytokeratin 15 - m
11	64	83.1	456	1 KRH05	keratin 15, type I
12	64	83.1	472	1 KRH0E	keratin 14, type I
13	64	83.1	473	2 JC4313	keratin 16, type I
14	64	83.1	486	1 KRXL	keratin 3, type I,
15	64	83.1	526	1 KRBOVI	keratin, 54K type
16	64	83.1	561	2 A31994	keratin 10, type I
17	64	83.1	593	1 KRH0U	keratin 10, type I
18	62	80.5	599	1 KRMBE1	keratin, 59K type
19	61	79.2	430	2 S05481	keratin 18, type I
20	60	77.9	399	2 A25470	cytokeratin 19 - b
21	60	77.9	419	2 A25438	keratin, type I cy
22	60	77.9	420	2 A37343	keratin 13, type I
23	60	77.9	458	1 KRH03	keratin 13, type-I
24	59	76.6	368	2 A28825	keratin, type I no
25	59	76.6	473	4 A33652	probable keratin 1
26	58	75.3	433	2 S01631	keratin, type I, e
27	58	75.3	467	2 I50476	keratin type I - g
28	58	75.3	570	2 S07330	keratin, epidermal
29	57	74.0	362	2 I37459	keratin Ha3-II, ty

#### ALIGNMENTS

##### RESULT 1

I37984  
keratin 9, type I, cytoskeletal - human  
N:Alternate names: cytokeratin 9; scatter protein 60K chain, placental  
C:Species: Homo sapiens (man)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text change 09-Jul-2004  
C:Accession: I37984; S7921; S41161; B35494; I37943  
R:Reis, A.; Hennies, H.C.; Langbein, L.; Digweed, M.; Mischke, D.; Drechsler, M.; Schroet  
Nature Genet. 6, 174-179, 1994  
A:Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).  
A:Reference number: I37984; MUID:94214498; PMID:7512862  
A:Accession: I37984  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-622 <RES>  
A:Cross-references: UNIPROT:P35527; UNIPARC:UPI000012DAE1; EMBL:X75015; NID:9453154; PIDN  
R:Langbein, L.; Heid, H.W.; Moll, I.; Franke, W.W.  
Differentiation 55, 57-71, 1993  
A:Title: Molecular characterization of the body site-specific human epidermal cytokeratin  
A:Reference number: I37943; MUID:94131202; PMID:7507869  
A:Accession: S40307  
A:Molecule type: mRNA  
A:Residues: 1-622 <LAN>  
A:Cross-references: UNIPARC:UPI000012DAE1; EMBL:Z29074; NID:9435475; PIDN:CAA82315.1; PII  
A:Accession: S77921  
A:Molecule type: protein  
A:Residues: 29-53; 62-104; 167-188; 199-233; 241-249; 295-362; 374-430; 450-480; 579-604 <LAF>  
A:Cross-references: UNIPARC:UPI00001774A9; UNIPARC:UPI00001774AA; UNIPARC:UPI00001774AB;  
4B0; UNIPARC:UPI00001774B1  
R:Langbein, L.  
submitted to the EMBL Data Library, December 1993  
A:Reference number: S41161  
A:Accession: S41161  
A:Molecule type: mRNA  
A:Residues: 1-11, 'SR', 13-622 <LAW>  
A:Cross-references: UNIPARC:UPI000016ABA2; EMBL:Z29074; NID:9435475; PIDN:CAA82315.1; PII  
R:Rosen, E.M.; Meromsky, L.; Romero, R.; Setter, E.; Goldberg, I.  
Biochem. Biophys. Res. Commun. 168, 1082-1088, 1990  
A:Title: Human placenta contains an epithelial scatter protein.  
A:Reference number: A35494; MUID:90267446; PMID:2140676  
A:Accession: B35494  
A:Molecule type: protein  
A:Residues: 'X', 450-465 <ROS>  
A:Cross-references: UNIPARC:UPI00001774B2  
C:Genetics:  
A:Gene: GDB:KRT9; EPPK  
A:Cross-references: GDB:303970; OMIM:144200  
A:Map position: 17q12-17q21  
A:Introns: 213/3; 241/2; 293/3; 347/3; 389/3; 464/2  
A>Note: defects in this gene may cause epidermolytic palmoplantar keratoderma  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; intermediate filament

30	57	74.0	404	2 JS0073	keratin, 47.6K typ
31	57	74.0	412	1 KRSHL1	keratin, 48K type
32	57	74.0	416	2 S60034	keratin Ha1, type
33	57	74.0	483	2 A55033	keratin 12 - mouse
34	55	71.4	392	2 A60777	keratin 2, type I,
35	55	71.4	407	2 I48739	MHa2(keratin acidi
36	55	71.4	416	2 A61404	keratin A, type I
37	55	71.4	416	2 A45559	keratin, type I, h
38	55	71.4	424	2 S37780	keratin 20, type I
39	55	71.4	429	2 A25145	keratin, 47K type
40	55	71.4	429	2 A40452	keratin 21, type I
41	53	68.8	455	2 A43950	vimentin - common
42	53	68.8	560	2 T16562	hypothetical prote
43	53	68.8	575	2 S48329	intermediate filam
44	53	68.8	576	2 S12277	intermediate filam
45	53	68.8	582	2 S24545	intermediate filam

```

F:1-153/Domain: head #status predicted <HEA>
F:154-459/Domain: helical rod #status predicted <ROD>
F:460-622/Domain: tail #status predicted <TAIL>

Query Match      100.0%; Score 77; DB 2; Length 622;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 STMQELNSRLASYLDK 16
DB      154 STMQELNSRLASYLDK 169

RESULT 2
I59463
N:Alternate names: endo B cytokeletin; keratin D
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence revision 02-Aug-1996 #text_change 09-Jul-2004
C:Accession: I59463; A25621; A28428; JT0406
R:Alonso, A.; Weber, T.; Jorcano, J.L.;
Roux's Arch. Dev. Biol. 196, 16-21, 1987
A:Title: Cloning and characterization of keratin D, a murine endodermal cytoskeletal protein
A:Reference number: I59463
A:Accession: I59463
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-423 <RES>
A:Cross-references: UNIPROT:P05784; UNIPARC:UPI000016CE78; GB:M36376; NID:G198587; PIDN:
R:Singer, P.A.; Trevor, K.; Oshima, R.G.
J. Biol. Chem. 261, 538-547, 1986
A:Title: Molecular cloning and characterization of the endo B cytokeletin expressed in P
A:Reference number: A25621; MUID:86085876; PMID:2416755
A:Accession: A25621
A:Molecule type: mRNA
A:Residues: 1-243, 'D', 245-252, 'A', 254-423 <SIN>
A:Cross-references: UNIPARC:UPI0000164A33; GB:M11686; NID:G198620; PIDN:AAA39390.1; PID:
R:Oshima, R.G.; Trevor, K.; Shevinsky, L.H.; Ryder, O.A.; Cecena, G.
Genes Dev. 2, 505-516, 1988
A:Title: Identification of the gene coding for the endo B murine cytokeletin and its med
A:Reference number: A28428; MUID:88255838; PMID:2454868
A:Accession: A28428
A:Molecule type: DNA
A:Residues: 1-132 <OSH>
A:Cross-references: UNIPARC:UPI000011E072; GB:Y00217; NID:G50842; PIDN:CAA68365.1; PID:G
R:Ichinose, Y.; Morita, T.; Zhang, F.; Srimahasonggram, S.; Tondella, M.L.C.; Matsumoto,
Gene 70, 85-95, 1988
A:Title: Nucleotide sequence and structure of the mouse cytokeletin endoB gene.
A:Reference number: JT0406; MUID:89196920; PMID:2467843
A:Accession: JT0406
A:Molecule type: DNA
A:Residues: 1-133, 'F', 135-243, 'D', 245-252, 'A', 254-423 <ICH>
A:Cross-references: UNIPARC:UPI000016CD33; GB:M22832; NID:G340757; PIDN:AAA37552.1; PID:
C:Genetics:
A:Gene: endoB; KERG
A:Introns: 132/3; 160/2; 212/3; 267/3; 309/3; 384/2
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament

Query Match      85.7%; Score 66; DB 2; Length 423;
Best Local Similarity 86.7%; Pred. No. 0.001;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 TMQELNSRLASYLDK 16
DB      75 TMQELNSRLASYLDK 89

RESULT 3
S30433
N:Alternate names: cytoskeletal - human
C:Species: Homo sapiens (man)

```

```

C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: S30433; S23648; I78538; B29139
R:Troyanovsky, S.W.; Leube, R.E.; Franke, W.W.
Eur. J. Cell Biol. 59, 127-137, 1992
A:Title: Characterization of the human gene encoding cytokeletin 17 and its expression p
A:Reference number: S30433; MUID:93105967; PMID:1281771
A:Accession: S30433
A:Molecule type: DNA
A:Residues: 1-432 <TRO>
A:Cross-references: UNIPROT:Q04695; UNIPARC:UPI0000148FD6; EMBL:X62571; NID:G30378; PIDN:
R:Flohr, T.; Buwitt, U.; Bonnekoh, B.; Decker, T.; Boettger, E.C.
Eur. J. Immunol. 22, 975-979, 1992
A:Title: Interferon-gamma regulates expression of a novel keratin class I gene.
A:Reference number: S23648; MUID:92201318; PMID:1372562
A:Accession: S23648
A:Molecule type: mRNA
A:Residues: 1-432 <FLO>
A:Cross-references: UNIPARC:UPI0000148FD6; EMBL:X62571; NID:G34074; PIDN:CAA44451.1; PID:
R:McLean, W.H.I.; Rugg, E.L.; Lunny, D.P.; Morley, S.M.; Lane, E.B.; Swensson, O.; Doppir
Kunkeler, L.; Munro, C.S.
Nature Genet. 9, 273-278, 1995
A:Title: Keratin 16 and keratin 17 mutations cause pachyonychia congenita.
A:Reference number: I58129; MUID:95291318; PMID:7539673
A:Accession: I78538
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 85-91, 'D', 93-101 <MCL>
A:Cross-references: UNIPARC:UPI000016B490; GB:S78515; NID:G1000378; PIDN:AAB34565.1; PID:
R:Kartasova, T.; Corneliussen, B.J.C.; Belt, P.; van de Putte, P.
Nucleic Acids Res. 15, 5945-5962, 1987
A:Title: Effects of UV, 4-NOO and TPA on gene expression in cultured human epidermal ker
A:Reference number: A29139; MUID:87316861; PMID:2442723
A:Accession: B29139
A:Molecule type: mRNA
A:Residues: 167, 'E', 169-175, 'D', 177-432 <KAR>
A:Cross-references: UNIPARC:UPI0000072C1D; GB:X05803; NID:G34080; PIDN:CAA29248.1; PID:G
C:Genetics:
A:Gene: GDB:KRT17
A:Cross-references: GDB:I136211; OMIM:148069
A:Map position: 17q12-17q21
A:Introns: 144/3; 172/2; 224/3; 278/3; 320/3; 394/2; 402/1
A:Note: defects in this gene may result in Jackson-Lawler pachyonychia congenita
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil

Query Match      84.4%; Score 65; DB 2; Length 432;
Best Local Similarity 81.2%; Pred. No. 0.0016;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 STMQELNSRLASYLDK 16
DB      86 ATMQNLNDRASYLDK 101

RESULT 4
S04511
keratin 3, type I, cytoskeletal (clone pUF451) - African clawed frog (fragment)
N:Alternate names: cytokeletin, 51K
C:Species: Xenopus laevis (African clawed frog)
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 10-Dec-1999
C:Accession: S04511
R:Hoffmann, W.; Sterrer, S.; Koenigstorfer, A.
FEBS Lett. 237, 178-182, 1988
A:Title: Amino acid sequence microheterogeneities of a type I cytokeletin of M(r) 51000
A:Reference number: S01327; MUID:89005622; PMID:2458965
A:Accession: S04511
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-327 <HOF>
A:Cross-references: UNIPARC:UPI00001774BA
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-310/Domain: helical rod #status predicted <ROD>

```







A>Note: defects in this gene may result in epidermolysis bullosa simplex, Dowling-Mearns C;Complex: heterotrimer of two type I and two type II proteins, usually keratin 5 (see C;Keywords: coiled coil; heterotrimer; intermediate filament F;2-472/Product: keratin, 50K type I cytoskeletal #status predicted <MAT> F;2-115/Domain: head <HD> F;2-115/Region: E1 and V1 subdomains F;116-426/Domain: rod <ROD> F;116-150/Region: coil 1A F;151-161/Region: linker 1 F;162-262/Region: coil 1B F;263-278/Region: linker 12 F;279-297/Region: coil 2A F;298-305/Region: linker 2 F;306-426/Region: coil 2B F;364/Region: stutter F;427-472/Domain: tail <END> F;427-472/Region: V2 and E2 subdomains

Query Match 83.1%; Score 64; DB 1; Length 472;  
Best Local Similarity 86.7%; Pred.No. 0.0026;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16  
||| ||| ||| |||  
DB 118 TMQNLDRLASYLDK 132

RESULT 13

JC4313

N:keratin 16, type I, cytoskeletal - human  
N:Alternate names: 46K keratin type I; cytokeratin 16  
C:Species: Homo sapiens (man)  
C>Date: 08-Dec-1995 #sequence\_revision 08-Feb-1996 #text\_change 09-Jul-2004  
C:Accession: JC4313; A24843; I58129  
R:Paladini, R.D.; Takahashi, K.; Gant, T.M.; Coulombe, P.A.  
Biochem. Biophys. Res. Commun. 215, 517-523, 1995  
A:Title: cDNA cloning and bacterial expression of the human type I keratin 16.  
A:Reference number: JC4313; MUID:96011809; PMID:7487986  
A:Accession: JC4313  
A:Molecule type: mRNA  
A:Residues: 1-473 <PAL>  
A:Cross-references: UNIPROT:P08779; UNIPARC:UPI000016B4A5; GB:S79867; NID:g195530; PIDN:  
A:Experimental source: epidermal keratinocytes  
R:Raychaudhuri, A.; Marchuk, D.; Lindhurst, M.; Fuchs, E.  
Mol. Cell. Biol. 6, 539-548, 1986  
A:Title: Three tightly linked genes encoding human type I keratins: conservation of sequence  
A:Reference number: A24843; MUID:87064338; PMID:2431270  
A:Accession: A24843  
A:Molecule type: DNA  
A:Residues: 1,'T',3-25,'A',27-37,'A',39-40,43,'ASTY',48-49,'A',51-186,'HAL',190-207,'ARPC'  
A:Cross-references: UNIPARC:UPI00001774B6; GB:M28439; NID:g186683; PIDN:AAS5460.1; PID:  
A>Note: The translated sequence in GenBank entry HUNKER16A8, release 111.0, (PIDN:AAA5944  
R:McLean, W.H.I.; Rugg, E.L.; Lunny, D.P.; Morley, S.M.; Lane, E.B.; Swenson, O.; Doppir  
Kunkeler, L.; Munro, C.S.  
Nature Genet. 9, 273-278, 1995  
A:Title: Keratin 16 and keratin 17 mutations cause pachyonychia congenita.  
A:Reference number: I58129; MUID:95291318; PMID:7539673  
A:Accession: I58129  
A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 118-131,'P',133-134 <MCL>  
A:Cross-references: UNIPARC:UPI000016B48F; GB:S78514; NID:g1000376; PIDN:AAB34564.1; PID  
A>Note: This is a mutant sequence  
C:Comment: This protein is an intermediate filament protein and expressed in epithelial t  
ound healing, psoriasis and cancer.  
C:Genetics:  
A:Gene: GDB:KRT16  
A:Cross-references: GDB:136207; OMIM:148067  
A:Map position: 17pter-17qter  
A>Note: Defects in this gene may result in Jadassohn-Lewandowsky pachyonychia congenita  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil

```

Query Match      83.1%; Score 64; DB 2; Length 473;
Best Local Similarity 86.7%; Pred. No. 0.0027;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TMQELNSRLASYLDK 16
Db 120 TMQNLNDRSLASYLDK 134

RESULT 14
KXKL
keratin 3, type I, cytoskeletal (clone pUF1001) - African clawed frog
N:Alternate names: cyokeratin, 51K
C:Species: Xenopus laevis (African clawed frog)
C>Date: 28-Aug-1985 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C:Accession: S01327; A02943
R:Hoffmann, W.; Sterrer, S.; Koenigstorfer, A.
FEBS Lett. 237, 178-182, 1988
A:Title: Amino acid sequence microheterogeneities of a type I cyokeratin of M(r) 51000
A:Reference number: S01327; PMID:89005622; PMID:2458965
A:Accession: S01327
A:Molecule type: mRNA
A:Residues: 1-486 <HOF>
A:Cross-references: UNIPROT:P02537; UNIPARC:UPI000012DADB; EMBL:X00968; NID:g64490; PIDN:
R:Hoffmann, W.; Franz, J.K.
EMBO J. 3, 1301-1306, 1984
A:Title: Amino acid sequence of the carboxy-terminal part of an acidic type I cyokerati
A:Reference number: A02943; PMID:84261417; PMID:6204859
A:Accession: A02943
A:Molecule type: mRNA
A:Residues: 339-351, 'T', 353-486 <HOF2>
A:Cross-references: UNIPARC:UPI00000FC7EB; GB:X00629; NID:g64868; PIDN:CAA25263.1; PID:9
A:Note: this sequence has been revised in paper S01327
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-120/Domain: head <HEA>
F:121-438/Domain: helical rod #status predicted <ROD>
F:439-486/Domain: tail <TAI>

Query Match      83.1%; Score 64; DB 1; Length 486;
Best Local Similarity 86.7%; Pred. No. 0.0027;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TMQELNSRLASYLDK 16
Db 129 TMQNLNDRSLASYLDK 143

RESULT 15
KRB0V1
keratin, 54K type I cytoskeletal - bovine
N:Alternate names: 54-kDa type I keratin; cyokeratin VTB
C:Species: Bos primigenius taurus (cattle)
C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 09-Jul-2004
C:Accession: A02941; S07262
R:Rieger, M.; Jorcano, J.L.; Franke, W.W.
EMBO J. 4, 2261-2267, 1985
A:Title: Complete sequence of a bovine type I cyokeratin gene: conserved and variable i
A:Reference number: A02941; PMID:86081734; PMID:2416562
A:Accession: A02941
A:Molecule type: DNA
A:Residues: 1-526 <RIE>
A:Cross-references: UNIPROT:P06394; UNIPARC:UPI000012DAB2; GB:X02870; NID:g478; PIDN:CAA
R:Jorcano, J.L.; Rieger, M.; Franz, J.K.; Schiller, D.L.; Moll, R.; Franke, W.W.
J. Mol. Biol. 179, 257-281, 1984
A:Title: Identification of two types of keratin polypeptides within the acidic cyokerat
A:Reference number: S07262; PMID:85058191; PMID:6209405
A:Accession: S07262
A:Molecule type: mRNA
A:Residues: 281-466, 'PAAATAAEVORPRRFRROVG', 490-491, 494, 'PVAVARRRK', 504-526 <JOR>
A:Cross-references: UNIPARC:UPI000017JD56; EMBL:X01460
A:Note: this sequence has been revised in reference A02941
C:Genetics:

```

```

A:Gene: cyokeratin VTB
A:Introns: 190/3; 218/2; 270/3; 324/3; 366/3; 439/2; 525/2
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-127/Domain: head <HED>
F:1-127/Region: E1 and V1 subdomains
F:128-441/Domain: rod <ROD>
F:128-162/Region: coil 1A
F:163-176/Region: linker 1
F:177-277/Region: coil 1B
F:278-293/Region: linker 12
F:294-312/Region: coil 2A
F:313-320/Region: linker 2
F:321-441/Region: coil 2B
F:379/Region: stutter
F:442-526/Domain: tail <END>
F:442-526/Region: V2 and E2 subdomains

Query Match      83.1%; Score 64; DB 1; Length 526;
Best Local Similarity 86.7%; Pred. No. 0.003;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TMQELNSRLASYLDK 16
Db 130 TMQNLNDRSLASYLDK 144

Search completed: July 5, 2006, 19:20:25
Job time : 19.4175 secs

```

GenCore version 5.1.1.9

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:56 ; Search time 163.728 Seconds  
(without alignments)  
90.395 Million cell updates/sec

Title: US-10-774-928A-14

Perfect score: 77  
Sequence: 1 STMQELNRLASYLDK 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 7.2.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	100.0	623	1 K1C9 HUMAN	P35527 homo sapien
2	71	92.2	786	2 O18740 CANFA	O18740 canis faml
3	66	85.7	422	1 K1C18 MOUSE	P05784 mus musculu
4	66	85.7	422	1 K1C18 RAT	Q5bjy9 rattus norv
5	66	85.7	423	2 Q3TIX1 MOUSE	Q3tix1 mus musculu
6	66	85.7	423	2 Q3TJH6 MOUSE	Q3tjh6 m 15 days p
7	66	85.7	423	2 Q3TJW7 MOUSE	Q3tjw7 m blastocys
8	65	84.4	280	2 O97624 CANFA	O97624 canis faml
9	65	84.4	423	1 K1C19 CHICK	O93256 gallus gall
10	65	84.4	424	2 Q5K2N8 PROAT	Q5k2n8 prototerus
11	65	84.4	426	2 Q8N1P6 HUMAN	Q8nlp6 homo sapien
12	65	84.4	431	1 K1C17 HUMAN	O04695 homo sapien
13	65	84.4	432	1 K1C17 MOUSE	Q9qwl7 mus musculu
14	65	84.4	433	2 Q61F08 RAT	Q61f08 rattus norv
15	64	83.1	123	2 Q9Z253 RAT	Q9z253 rattus norv
16	64	83.1	338	2 Q5K2P3 PROAT	Q5k2p3 prototerus
17	64	83.1	400	1 K1C19 HUMAN	P08727 homo sapien
18	64	83.1	400	1 K1C19_PONPY	Q5r8s9 pongo pygma
19	64	83.1	401	1 K1C19_POTTR	P19001 mus musculu
20	64	83.1	403	1 K1C19_MOUSE	Q63279 rattus norv
21	64	83.1	403	1 K1C19 RAT	Q28706 oryctolagus
22	64	83.1	411	1 K1C12 RABIT	Q61rp5 xenopus lae
23	64	83.1	412	2 Q61RP5 XENLA	P08730 mus musculu
24	64	83.1	437	1 K1C13 MOUSE	Q61fv3 rattus norv
25	64	83.1	447	1 K1C15_MOUSE	Q61414 mus musculu
26	64	83.1	452	1 K1C15_MOUSE	P19012 homo sapien
27	64	83.1	456	1 K1C15 HUMAN	Q642p1 xenopus lae
28	64	83.1	467	1 K1C14 XENLA	P02533 homo sapien
29	64	83.1	471	1 K1C14 HUMAN	P08779 homo sapien
30	64	83.1	472	1 K1C16 HUMAN	Q50514 xenopus lae
31	64	83.1	483	2 Q50514 XENLA	

## RESULT 1

K1C9 HUMAN	64	83.1	486	1	K1C0 XENLA	P02537 xenopus lae
ID K1C9 HUMAN	64	83.1	491	2	Q8AV12 XENLA	Q8av12 xenopus lae
AC P35527; O00109; Q14665;	64	83.1	494	1	K1C12 HUMAN	Q99456 homo sapien
DT 01-JUN-1994, integrated into UniProtKB/Swiss-Prot.	64	83.1	526	1	K1C10_BOVIN	P06394 bos taurus
DT 08-NOV-2005, sequence version 2.	64	83.1	561	2	O14664 HUMAN	Q14664 homo sapien
DT 07-FEB-2006, entry version 50.	64	83.1	568	2	Q6EIZO CANFA	Q6eiz0 canis famil
DE Keratin, type I cytoskeletal 9 (Cytokeratin-9) (CK-9) (Keratin-9) (K9).	64	83.1	584	2	Q8N175 HUMAN	Q8n175 homo sapien
GN Name=KRT9;	64	83.1	593	1	K1C10_HUMAN	P13845 homo sapien
OS Homo sapiens (Human).	63	81.8	437	2	Q8BKC6 MOUSE	Q8bkc6 mus musculu
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	62	80.5	334	2	Q9JKB4 MOUSE	Q9jkb4 mus musculu
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;	62	80.5	414	2	Q98UJ2 XENLA	Q98uj2 xenopus lae
OC Homo.	62	80.5	431	2	O566F9 XENLA	O566f9 xenopus lae
OX NCBI_TaxID=9606;	62	80.5	431	2	Q6GLQ7 XENLA	Q6glq7 xenopus lae
RN [1]	62	80.5	456	2	Q5K2P1_PROAT	Q5k2p1 prototerus
RP NUCLEOTIDE SEQUENCE [MRNA], AND PARTIAL PROTEIN SEQUENCE.						
RC TISSUE=Foot sole tissue;						
RX MEDLINE=94131202; PubMed=7507869;						
RA Langbein L., Heid H.W., Moll I., Franke W.W.;						
RT "Molecular characterization of the body site-specific human epidermal						
RT cytokeratin 9: cDNA cloning, amino acid sequence, and tissue						
RT specificity of gene expression.";						
RT Differentiation 55:57-72(1993).						
RL [2]						
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANTS EPPK LYS-161; GLN-163						
RP AND TRP-163.						
RX MEDLINE=94214498; PubMed=7512862;						
RA Reis A., Hennies H.-C., Langbein L., Digweed M., Mischke D.,						
RA Dreschler M., Schroek E., Royer-Pokora B., Franke W.W., Sperling K.,						
RA Kuester W.;						
RT "Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma						
RT (EPPK).";						
RL Nat. Genet. 6:174-179(1994).						
RN [3]						
RP PROTEIN SEQUENCE OF 14-29, AND MASS SPECTROMETRY.						
RC TISSUE=cervix carcinoma;						
RA Bienvenut W.V.;						
RL Unpublished observations (AUG-2005).						
RN [4]						
RP NUCLEOTIDE SEQUENCE [MRNA] OF 147-372, AND VARIANT EPPK GLN-163.						
RX MEDLINE=96228052; PubMed=8647270; DOI=10.1016/0014-5793(96)00393-6;						
RA Kobayashi S., Tanaka T., Matsuyoshi N., Imamura S.;						
RT "Keratin 9 point mutation in the pedigree of epidermolytic hereditary						
RT palmoplantar keratoderma perturbs keratin intermediate filament						
RT network formation.";						
RL FEBS Lett. 386:149-155(1996).						
RN [5]						
RP PROTEIN SEQUENCE OF 450-466.						
RX MEDLINE=90267446; PubMed=2140676;						
RA Rosen E.M., Meromsky L., Romero R., Setter E., Goldberg I.;						
RT "Human placenta contains an epithelial scatter protein.";						
RL Biochem. Biophys. Res. Commun. 168:1082-1088(1990).						
RN [6]						



OX NCBI\_TaxID=9615;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Lacharme P., Hitt C., Jouquand S., Priat C., Galibert F.;  
 RT "Identification and analysis of the dog keratin 9 (KRT9) gene.";  
 RL Anim. Genet. 9:173-178(1998).  
 CC  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC ENBL; AF000949; AAC26971.1; -; Genomic\_DNA.  
 DR HSSP; P08670; 1GK7.  
 DR Ensembl; ENSCAF00000015952; Canis familiaris.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR InterPro; IPR002952; Eggshell.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin\_I.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01228; EGGSHLL.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 SQ SEQUENCE 786 AA; 76354 MW; 272AB5425DD09535 CRC64;  
  
 Query Match 92.2%; Score 71; DB 2; Length 786;  
 Best Local Similarity 87.5%; Pred. No. 0.004;  
 Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 STWQELNSRLASYLDK 16  
 Db 139 TTMQDLNRLASYLDK 154  
  
 RESULT 3  
 KIC18\_MOUSE  
 ID KIC18\_MOUSE STANDARD; PRT; 422 AA.  
 AC P05784; Q61766;  
 DT 01-NOV-1988, integrated into UniProtKB/Swiss-Prot.  
 DT 15-JUL-1998, sequence version 51.  
 DT 07-FEB-2006, entry version 51.  
 DE Keratin, type I cytoskeletal 18 (Cytokeratin-18) (CK-18) (Keratin-18)  
 DE (K18) (Cytokeratin endo B) (Keratin D).  
 GN Name=Krt18; Synonyms=Kerd, Krt1-18;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=89196920; PubMed=2467843; DOI=10.1016/0378-1119(88)90107-2;  
 RA Ichinose Y., Morita T., Zhang F., Srinivasanongram S., Tondella M.L.C.,  
 RA Matsumoto M., Nozaki M., Matsushiro A.;  
 RT "Nucleotide sequence and structure of the mouse cytoke-  
 RT ratin endo B gene.";  
 RL Gene 70:85-95(1988).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Teratocarcinoma;  
 RA Alonso A., Weber T., Jorcano J.L.;  
 RT "Cloning and characterization of keratin D, a murine endodermal  
 RT cytoskeletal protein induced during in vitro differentiation of F9  
 RT teratocarcinoma cells.";  
 RL Roux's Arch. Dev. Biol. 196:16-21(1987).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=86085876; PubMed=2416755;  
 RA Singer P.A., Trevor K., Oshima R.G.;  
 RT "Molecular cloning and characterization of the Endo B cytoke-  
 RT ratin expressed in preimplantation mouse embryos.";  
 RL J. Biol. Chem. 261:538-547(1986).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE OF 1-131.  
 RX MEDLINE=88255838; PubMed=2454868;  
 RA Oshima R.G., Trevor K., Shevinsky L.H., Ryder O.A., Cecena G.;

RT "Identification of the gene coding for the Endo B murine cytoke-  
 RT ratin and its methylated, stable inactive state in mouse non-epithelial  
 RL cells.";  
 RL Genes Dev. 2:505-516(1988).  
 CC -1- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
 CC Keratin 18 associates with keratin 8. Interacts with PNN (By  
 CC similarity).  
 CC -1- MISCELLANEOUS: There are two types of cytoskeletal and  
 CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
 CC 55 and 56-70 kDa, respectively).  
 CC -1- SIMILARITY: Belongs to the intermediate filament family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC ENBL; M22832; AAA37552.1; -; Genomic\_DNA.  
 DR EMBL; M36376; AAA39373.1; -; mRNA.  
 DR EMBL; M11686; AAA39390.1; -; mRNA.  
 DR EMBL; Y00217; CAA68365.1; -; Genomic\_DNA.  
 DR PIR; I59463; I59463.  
 DR HSSP; P08670; 1GK7.  
 DR SWISS-2DPAGE; P05784; MOUSE.  
 DR Ensembl; ENSMUSG00000023043; Mus musculus.  
 DR MGI; MGI:96692; Krt1-18.  
 DR GO; GO:0005739; C: mitochondrion; IDA.  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR002957; Keratin\_I.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPE1KERATIN.  
 DR PROSITE; PS00226; IF; 1  
 KW Acetylation; Coiled coil; Glycoprotein; Intermediate filament;  
 KW Keratin.  
 FT INIT MET 0 0 By similarity.  
 FT CHAIN 1 422 Keratin, type I cytoskeletal 18.  
 FT /FTID=PRO\_0000063667.  
 FT REGION 1 70 Head.  
 FT REGION 61 365 Necessary for interaction with PNN (By  
 FT similarity).  
 FT REGION 71 379 Rod.  
 FT REGION 71 106 Coiled coil 1A.  
 FT REGION 107 124 Linker 1.  
 FT REGION 125 216 Coiled coil 1B.  
 FT REGION 217 240 Linker 12.  
 FT REGION 241 379 Coiled coil 2.  
 FT REGION 380 422 Tail.  
 FT SITE 263 263 Stutter.  
 FT SITE 323 323 Stutter.  
 FT MOD RES 1 1 N-acetylserine (By similarity).  
 FT CARBOHYD 30 30 O-linked (GlcNAc) (By similarity).  
 FT CARBOHYD 31 31 O-linked (GlcNAc) (By similarity).  
 FT CARBOHYD 49 49 O-linked (GlcNAc) (By similarity).  
 FT CONFLICT 133 133 L -> F (in Ref. 1).  
 FT CONFLICT 243 243 D -> N (in Ref. 2).  
 FT CONFLICT 252 252 A -> G (in Ref. 2).  
 SQ SEQUENCE 422 AA; 47373 MW; 4DSB0E9C7732F2F3 CRC64;  
  
 Query Match 85.7%; Score 66; DB 1; Length 422;  
 Best Local Similarity 86.7%; Pred. No. 0.015;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 2 TWQELNSRLASYLDK 16  
 Db 74 TTMQDLNRLASYLDK 88  
  
 RESULT 4  
 KIC18\_RAT  
 ID KIC18\_RAT STANDARD; PRT; 422 AA.  
 AC Q5BJY9; Q63278;  
 DT 16-AUG-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 16-AUG-2005, sequence version 2.  
 DT 07-MAR-2006, entry version 13.  
 DE Keratin, type I cytoskeletal 18 (Cytokeratin-18) (CK-18) (Keratin-18)

```
DE GN Name=Krt18; Synonyms=Krt1-18;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Thymus;
RG NIH - Mammalian Gene Collection (MGC) project;
RL Submitted (MAR-2005) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] OF 123-269, AND DEVELOPMENTAL STAGE.
RC STRAIN=Wistar; TISSUE=Fetal gonad;
RX MEDLINE=96081364; PubMed=8541209; DOI=10.1016/0925-4773(95)00401-L;
RA Fridmacher V., le Bert M., Guillou F., Magre S.;
RT "Switch in the expression of the K19/K18 keratin genes as a very early
RL evidence of testicular differentiation in the rat.";
RN Mech. Dev. 52:199-207(1995).
RN [3]
RP PROTEIN SEQUENCE OF 130-139, AND SUBCELLULAR LOCATION.
RC TISSUE=Liver;
RX Segawa M., Niino K., Mineki R., Kaga N., Murayama K., Sugimoto K.,
RA Watanabe Y., Furukawa K., Horigome T.;
RT "Proteome analysis of a rat liver nuclear insoluble protein fraction
RT and localization of a novel protein, ISP36, to compartments in the
RT interchromatin space.";
RL FEBS J. 272:4327-4338(2005).
CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.
CC Keratin 18 associates with keratin 8. Interacts with PNN (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear matrix.
CC -!- DEVELOPMENTAL STAGE: Expression first detected in the male gonad
CC at 13.5 gpc, at the onset of testicular differentiation.
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and
CC microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
CC 55 and 56-70 kDa, respectively).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC -----
DR EMBL; BC091275; AAH91275.1; -; mRNA.
DR EMBL; X81448; CAA57204.1; -; mRNA.
DR Ensembl; ENSRN00000022699; Rattus norvegicus.
DR RGD; 619935; Krt1-18.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin I.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF, 1
KW Acetylation; Coiled coil; Direct protein sequencing; Glycoprotein;
KW Intermediate filament; Keratin; Nuclear protein.
FT INIT MET 0 0 By similarity.
FT CHAIN 1 422 Keratin, type I cytoskeletal 18.
FT /FTID=PRO_0000063668.
FT REGION 1 70 Head
FT REGION 61 365 Necessary for interaction with PNN (By
FT similarity).
FT REGION 71 379 Rod.
FT REGION 71 106 Coiled coil 1A.
FT REGION 107 124 Linker 1.
FT REGION 125 216 Coiled coil 1B.
FT REGION 217 240 Linker 12.
FT REGION 241 379 Coiled coil 2.
FT REGION 380 422 Tail.
FT SITE 263 263 Stutter.
FT SITE 323 323 Stutter.
FT MOD_RES 1 1 N-acetylserine (By similarity).
FT CARBOHYD 30 30 O-linked (GlcNAc) (By similarity).
FT CARBOHYD 31 31 O-linked (GlcNAc) (By similarity).
FT ~FT

FT CARBOHYD 49 49 O-linked (GlcNAc) (By similarity).
FT CONFLICT 124 124 T -> I (in Ref. 2).
SQ SEQUENCE 422 AA; 47630 MW; CB7D7DD31FAG3FE7 CRC64;

Query Match 85.7%; Score 66; DB 1; Length 422;
Best Local Similarity 86.7%; Pred. No. 0.015; Indels 0; Gaps 0;
Matches 13; Conservative 1; Mismatches 1;

QY 2 TMOELNSRLASYLDK 16
|||:|||||
DB 74 TMOELNDELASYLDK 88

RESULT 5
Q3TIX1_MOUSE PRELIMINARY; PRT; 423 AA.
AC Q3TIX1_
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE 14 days pregnant adult female placenta cDNA, RIKEN full-length
DE enriched library, clone: I530028D16 product: keratin complex 1, acidic,
DE gene 18, full insert sequence.
GN Name=Krt1-18;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Placenta;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Placenta;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impiombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humnick L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelsey J., Kitamura H.,
RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugiu K., Sultana R., Takenaka Y., Taki K.,
RA Tannoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
```

RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Placenta;  
RX PubMed=16141073; DOI=10.1126/science.11112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RT (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Placenta;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Kogut D.R., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Mulcais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sadelain A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wyszewski-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shingawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Placenta;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleschmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bernaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schonbach C., Seva T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Placenta;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Placenta;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Placenta;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Horii F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted under the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AK167676; BAE39725.1; -; mRNA.  
DR MGI: MGI:96692; Krt1-18.  
DR GO: GO:0005739; C:mitochondrion; IDA.  
DR InterPro: IPR001664; IF.  
DR InterPro: IPR002957; Keratin\_1.  
DR Pfam: PF00038; Filament; 1.  
DR PRINTS: PR01248; TYPE1KERATIN.  
DR PROSITE: PS00226; IF; 1.  
KW Intermediate filament; Keratin.  
SQ SEQUENCE 423 AA; 47466 MW; A679AD6974A22118 CRC64;  
Query Match 85.7%; Score 66; DB 2; Length 423;  
Best Local Similarity 86.7%; Pred. No. 0.015;  
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TMQELNRLASYLDK 16  
DB 75 TMQELNRLASYLDK 89  
|||||  
RESULT 6  
Q3TJH6 MOUSE PRELIMINARY; PRT; 423 AA.  
ID Q3TJH6; Q3TJH6; Q3TJH6;  
AC Q3TJH6;  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 11-OCT-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE 15 days pregnant adult female placenta cDNA, RIKEN full-length  
DE enriched library, clone:1530014C23 product:keratin complex 1, acidic,  
DE gene 18, full insert sequence (1530014C23 product:keratin complex 1, acidic,  
DE cDNA, RIKEN full-length enriched library, clone:1530014C20  
DE product:keratin complex 1, acidic, gene 18, full insert sequence)  
DE (Blastocyst blastocyst cDNA, RIKEN full-length enriched library,  
DE clone:1530014C23 product:keratin complex 1, acidic, gene 18, full  
DE insert sequence).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridea; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impiombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matveeva H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamagishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Hayashizaki Y.;  
 RA Matick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragni T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gisel C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shirazee-Kishikawa T., Konno H., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RT Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Mateu Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 RA Blake J., Bonfelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Watahiki K.,  
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Iehikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Placenta;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Horii F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,



```

RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
RA Muranatsu M., Hayashizaki Y.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK167432; BAE39519.1; -; mRNA.
DR EMBL; AK167469; BAE39553.1; -; mRNA.
DR EMBL; AK167072; BAE39232.1; -; mRNA.
DR GO; GO:0005739; C-mitochondrion; IDA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin I.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 423 AA; 47538 MW; A67ACD6814A02118 CRC64;

Query Match 85.7%; Score 66; DB 2; Length 423;
Best Local Similarity 86.7%; Pred. No. 0.015;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 2 TMOQLNSRLASYLDK 16
Db 75 TMOQLNDRLASYLDK 89

RESULT 7
Q3TJW7 MOUSE PRELIMINARY; PRT; 423 AA.
AC Q3TJW7;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 5.
DE Blastocyst blastocyst cDNA, RIKEN full-length enriched library,
DE clone:11C0047008 product:keratin complex 1, acidic, gene 18, full
DE insert sequence (4 days embryo whole body cDNA, RIKEN full-length
DE enriched library, clone:10C0039P11 product:keratin complex 1, acidic,
DE gene 18, full insert sequence).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
CA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX PubMed=16141072; DOI=10.1126/science.1112014;
CA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Adams J., Allen J.E.,
RA Davis M.J., Wilming L.G., Adkins V., Chabot R., Scheraga H.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal K.P., Baxter L., Belsel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Chothia C., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,

```

```

RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rest B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Gironde S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Lida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX PubMed=16141073; DOI=10.1126/science.1112009;
RG RIKEN Genome Exploration Research Group, and Genome Science Group
RG (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriell L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Gough J.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayashizaki Y.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Harada A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=21085660; PubMed=1217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

```

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Kasavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Mateu Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guscinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzei J., Mombarts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body; DOI=10.1101/gr.145100;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata K., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body; DOI=10.1101/gr.152600;  
RX MEDLINE=20530913; PubMed=11076961; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK167265; BAE39378.1; -; mRNA.  
DR EMBL; AK145413; BAE26424.1; -; mRNA.  
DR GO; GO:0005739; C:mitochondrion; IDA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.

KW Intermediate filament; Keratin.  
SQ SEQUENCE 423 AA; 47510 MW; C70574E694E89898 CRC64;  
Query Match 85.7%; Score 66; DB 2; Length 423;  
Best Local Similarity 86.7%; Pred. No. 0.015;  
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TMQELNSRLASYLDK 16  
:|||||  
DB 75 TMQDLNDRLASLYDK 89  
:|||||  
RESULT 8  
O97624 CANFA PRELIMINARY; PRT; 280 AA.  
ID O97624 CANFA PRELIMINARY; PRT; 280 AA.  
AC O97624;  
DT 01-MAY-1999, integrated into UniProtKB/TrEMBL.  
DT 01-MAY-1999, sequence version 1.  
DT 07-FEB-2006, entry version 23.  
DE Keratin 17 (Fragment).  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
OC Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=99189141; PubMed=10087295; DOI=10.1007/s003359901004;  
RA Miller A.B., Breen M., Murphy K.E.;  
RT "Chromosomal localization of acidic and basic keratin genes of the  
RT domestic dog.";  
RL Mamm. Genome 10:371-375(1999).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AF084571; AAD11416.1; -; Genomic\_DNA.  
DR HSSP; P08670; 1GK7.  
DR Ensembl; ENSCAFG00000015944; Canis familiaris.  
DR LinkHub; O97624;  
DR GO; GO:0005882; C:intermediate filament; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR Pfam; PF00038; Filament; 1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
KW Keratin.  
FT NON TER 280 280  
SQ SEQUENCE 280 AA; 30557 MW; 2758FBE80E31B04D CRC64;  
Query Match 84.4%; Score 65; DB 2; Length 280;  
Best Local Similarity 81.2%; Pred. No. 0.015;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 STMQELNSRLASYLDK 16  
:|||||  
DB 86 ATMQNLNDRLASLYDK 101  
:|||||  
RESULT 9  
K1C19 CHICK STANDARD; PRT; 423 AA.  
ID K1C19 CHICK STANDARD; PRT; 423 AA.  
AC O93256;  
DT 07-MAR-2006, integrated into UniProtKB/Swiss-Prot.  
DT 01-NOV-1998, sequence version 1.  
DT 07-MAR-2006, entry version 27.  
DE Keratin, type I cytoskeletal 19 (Cytokeratin-19) (CK-19) (Keratin-19)  
DE (K19) (GK-19).  
GN Name=KRT19;  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.

OX NCBI\_TaxID=9031;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND DEVELOPMENTAL STAGE.  
RC TISSUE=Embryo;  
RX MEDLINE=98152854; PubMed=9493835;  
RX DOI=10.1046/j.1440-169X.1997.t01-5-00011.x;  
RA Sato K., Yasugi S.;  
RT "Chicken keratin-19: cloning of cDNA and analysis of expression in the  
Dev. Growth Differ. 39:751-761(1997).  
CC -!- FUNCTION: Involved in the organization of myofibers. Together with  
CC KRT8, helps to link the contractile apparatus to dystrophin at the  
CC costameres of striated muscle (By similarity).  
CC -!- SUBUNIT: Heterotrimer of two type I and two type II keratins.  
CC Interacts with PNN and the actin-binding domain of DMD (By  
CC similarity).  
CC -!- DEVELOPMENTAL STAGE: Expressed in the digestive tract throughout  
CC development in the epithelia of proventriculus and glandular  
CC structures. In 13 day embryo, strongly expressed in esophagus with  
CC moderate expression in proventriculus and lung. Weak expression in  
CC gizzard, small intestine, lung and dorsal skin. In embryos over 13  
CC days old, observed in ectodermal, endodermal epithelium and also  
CC in neural and mesodermal tissues (i.e. notochord), floorplate in  
CC the neural tube, somatic mesoderm, splanchnic mesoderm and  
CC dermatome. In more developed embryos, expression was localized in  
CC the anterior lobe of the pituitary gland, notochord and  
CC hypothalamus of the diencephalon (derived from the floor plate of  
CC the neural tube). Also localized in the mesonephric mesoderm and  
CC lateral plate mesoderm.  
CC -!- DOMAIN: This keratin differs from all other IF proteins in lacking  
CC the C-terminal tail domain.  
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and  
CC microfibrillar keratin, I (acidic; 40-55 kDa) [K9 to K20] and II  
CC (neutral to basic; 56-70 kDa) [K1 to K8].  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AB016281; BAA31952.1; -; mRNA.  
CC HSSP: P08670; 1GK7.  
DR Ensembl: ENSGALG0000003843; Gallus gallus.  
DR InterPro: IPR001664; IF.  
DR InterPro: IPR002957; Keratin\_I.  
DR Pfam: PF00038; Filament; 1.  
DR PRINTS: PR01248; TYPEIKERATIN.  
DR PROSITE: PS00226; IF; 1.  
KW Coiled coil; Intermediate filament; Keratin; Phosphorylation.  
FT CHAIN 1 423 Keratin, type I cytoskeletal 19.  
FT /Ftrid=PRO\_0000227527.  
FT REGION 1 93 Head.  
FT REGION 94 401 Rod.  
FT REGION 94 401 Coiled coil 1A.  
FT REGION 130 147 Linker 1.  
FT REGION 148 239 Coiled coil 1B.  
FT REGION 240 262 Linker 12.  
FT REGION 258 415 Necessary for interaction with PNN (By  
FT similarity).  
FT REGION 263 401 Coiled coil 2.  
FT REGION 402 414 Rod-like helical tail.  
FT COMPIAS 15 92 Gly-rich.  
FT SITE 281 281 Stutter.  
FT SITE 341 341 Stutter.  
FT MOD RES 33 33 Phosphoserine (By similarity).  
FT SEQUENCE 423 AA; 46082 MW; E03390A9558AD2D0 CRC64;  
Query Match 84.4%; Score 65; DB 1; Length 423;  
Best Local Similarity 81.2%; Pred. No. 0.023;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 STMOELNSRLASYLDK 16  
DB 96 TTQNLNDRSLASYLDK 111

RESULT 10  
Q5K2N8 PROAT PRELIMINARY; PRT; 424 AA.  
ID O5K2N8 PROAT PRELIMINARY; PRT; 424 AA.  
AC O5K2N8;  
DT 15-FEB-2005, integrated into UniProtKB/TrEMBL.  
DT 15-FEB-2005, sequence version 1.  
DT 07-FEB-2006, entry version 6.  
DE Type I keratin 19.  
GN Name=krt19;  
OS Procterus aethiopicus (Marbled lungfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Dipnoi; Lepidosireniiformes; Protopteridae; Protopterus.  
OX NCBI\_TaxID=7886;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Schaffeld M., Bremer M., Hunzinger C., Markl J.;  
RT "Evolution of tissue-specific keratins as deduced from novel cDNA  
RT sequences of the lungfish Protopterus aethiopicus.";  
RL Eur. J. Cell Biol. 84:363-377(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RA Bremer M.;  
RT "Primary structure and evolution of intermediate filament proteins  
RT from the lungfish Protopterus aethiopicus.";  
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AU785793; CAH05049.1; -; mRNA.  
CC GO: GO:0005882; C:intermediate filament; IEA.  
CC GO: GO:0005198; F:structural molecule activity; IEA.  
CC InterPro: IPR001664; IF.  
CC InterPro: IPR002957; Keratin\_I.  
CC InterPro: IPR002155; Thiolase.  
CC Pfam: PF00038; Filament; 1.  
CC PRINTS: PR01248; TYPEIKERATIN.  
CC PROSITE: PS00099; THIOLASE\_3; UNKNOWN\_1.  
KW Keratin.  
SQ SEQUENCE 424 AA; 46548 MW; C811F1B9E67E3419 CRC64;  
Query Match 84.4%; Score 65; DB 2; Length 424;  
Best Local Similarity 81.2%; Pred. No. 0.023;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 STMOELNSRLASYLDK 16  
DB 106 ATMQLNDRSLASYLDK 121

RESULT 11  
Q8NIP6 HUMAN PRELIMINARY; PRT; 426 AA.  
ID Q8NIP6 HUMAN PRELIMINARY; PRT; 426 AA.  
AC Q8NIP6;  
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2002, sequence version 1.  
DT 07-MAR-2006, entry version 19.  
DE CDNA FLJ38023 fis, clone CTONG2012901, highly similar to KERATIN, TYPE  
DE I CYTOSKELETAL 17.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OX Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA TISSUE=Tongue;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RX Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RX Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RX Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,

RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
RA Murakami K., Yasuda T., Iwavanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoi T., Katsu Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa B., Omura Y.,  
RA Abe K., Kamiyama K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tani H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fuji A., Hara H., Tanase T.-O., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Mueshahno K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RL Nat. Genet. 36:40-45(2004).  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC  
CC EMBL: AK095342; BAC04534.1; -; mRNA.  
DR HSSP: P08670; IKG7.  
DR Ensembl: ENSG00000186831; Homo sapiens.  
DR GO: GO:005882; C:intermediate filament; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR InterPro: IPR001664; IF.  
DR InterPro: IPR002957; Keratin\_I.  
DR Pfam: PF00038; Filament; 1.  
DR PRINTS: PR01248; TYPE1KERATIN.  
DR PROSITE: PS00226; IF; 1.  
KW Intermediate filament.  
SQ SEQUENCE 426 AA; 47603 MW; C423929F5260F437 CRC64;

Query Match 84.4%; Score 65; DB 2; Length 426;  
Best Local Similarity 81.2%; Pred. No. 0.023;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 STMQELNSRLASYLQK 16  
:||||| ||||| |||||  
Db 80 ATMQLNDRLASYLQK 95

RESULT 12  
K1C17 HUMAN  
ID K1C17 HUMAN STANDARD; PRT; 431 AA.  
AC Q04695;  
DT 01-JUN-1994, integrated into UniProtKB/Swiss-Prot.  
DT 01-JUN-1994, sequence version 1.  
DT 07-FEB-2006, entry version 52.  
DE Keratin, type I cytoskeletal 17 (Cytokeratin-17) (CK-17) (Keratin-17)  
DE (K17) (39.1).  
GN Name=KRT17;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=93105967; PubMed=1281771;

RA Troyanovsky S.M., Leube R.E., Franke W.W.;  
RT "Characterization of the human gene encoding cytokeratin 17 and its  
RT expression pattern.";  
RL Eur. J. Cell Biol. 59:127-137(1992).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=92201318; PubMed=1372562;  
RA Flohr T., Buitt U., Bonnekoh B., Decker T., Boettger E.C.;  
RT "Interferon-gamma regulates expression of a novel keratin class I  
RT gene.";  
RL Eur. J. Immunol. 22:975-979(1992).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RA Zimbelmann R.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Brain, Cervix, Lung, and Muscle;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.M., Hong F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP VARIANT PC2 ASP-91, AND NUCLEOTIDE SEQUENCE OF 84-100.  
RX MEDLINE=95291318; PubMed=7539673;  
RA McLean W.H.I., Rugg E.L., Lunny D.P., Morley S.M., Lane E.B.,  
RA Swenson O., Dopping-Hepenstal P.J.C., Griffiths W.A.D., Eady R.A.J.,  
RA Higgins C., Navearia H.A., Leigh I.M., Strachan T., Kunkeler L.,  
RA Munro C.S.;  
RA "Keratin 16 and keratin 17 mutations cause pachyonychia congenita";  
RL Nat. Genet. 9:273-278(1995).  
RN [6]  
RP VARIANTS PC2 SER-91 AND ASP-97, AND VARIANTS SM HIS-91 AND HIS-93.  
RX MEDLINE=97160918; PubMed=9008238; DOI=10.1111/1523-1747.ep12335315;  
RA Smth F.J.D., Corden L.D., Rugg E.L., Ratnavel R., Leigh I.M.,  
RA Moss C., Tidman M.J., Hohl D., Huber M., Kunkeler L., Munro C.S.,  
RA Lane E.B., McLean W.H.I.;  
RT "Missense mutations in keratin 17 cause either pachyonychia congenita  
RT type 2 or a phenotype resembling steatocystoma multiplex";  
RL J. Invest. Dermatol. 108:220-223(1997).  
RN [7]  
RP VARIANTS PC2 SER-91 AND CYS-93, AND VARIANT SM CYS-93.  
RX MEDLINE=99069010; PubMed=9767294;  
RX DOI=10.1046/j.1365-2133.1998.02413.x;  
RA Covello S.P., Smith F.J.D., Sillevius Smitt J.H., Faller A.S.,  
RA Munro C.S., Jonkman M.F., Uitto J., McLean W.H.I.;  
RT "Keratin 17 mutations cause either steatocystoma multiplex or  
RT pachyonychia congenita type 2";  
RL Br. J. Dermatol. 139:475-480(1998).  
RN [8]  
RP VARIANT PC2 THR-87.  
RX MEDLINE=20040456; PubMed=10571744;  
RX DOI=10.1046/j.1523-1747.1999.00762.x;  
RA Celebi J.T., Tanzi E.L., Yao Y.J., Michael E.J., Peacocke M.;  
RT "Mutation report: identification of a germline mutation in keratin 17  
RT in a family with pachyonychia congenita type 2";

RL J. Invest. Dermatol. 113:848-850(1999).  
 RN [9]  
 RP VARIANTS PC2 93-PRO--TYR-97 DEL; PRO-93 AND GLN-94.  
 RX MEDLINE=21246897; PubMed=11348474;  
 RA DOI=10.1046/j.1523-1747.2001.01335.x;  
 RA Smith F.J.D., Coleman C.M., Bayoumy N.M., Tenconi R., Nelson J.,  
 David A., McLean W.H.I.;  
 RT "Novel keratin 17 mutations in pachyonychia congenita type 2.";  
 RL J. Invest. Dermatol. 116:806-808(2001).  
 RN [10]  
 RP VARIANTS PC2 PRO-94; SER-96 DEL AND PRO-98.  
 RX MEDLINE=21884166; PubMed=11886499;  
 RA DOI=10.1046/j.0022-202x.2001.01565.x;  
 RA Terrinoni A., Smith F.J.D., Didona B., Canzona F., Paradisi M.,  
 Huber W., Mohl D., David A., Verloes A., Leigh I.M., Munro C.S.,  
 Melino G., McLean W.H.I.;  
 RT "Novel and recurrent mutations in the genes encoding keratins K6a, K16  
 and K17 in 13 cases of pachyonychia congenita.";  
 RL J. Invest. Dermatol. 117:1391-1396(2001).  
 RN [11]  
 RP VARIANTS PC2 MET-101.  
 RX MEDLINE=21864371; PubMed=11874497;  
 RA DOI=10.1046/j.0022-202x.2001.01701.x;  
 RA Hashiguchi T., Yocumoto S., Shimada H., Terasaki K., Setoyama M.,  
 Kobayashi K., Saheki T., Kanazaki T.;  
 RT "A novel point mutation in the keratin 17 gene in a Japanese case of  
 pachyonychia congenita type 2.";  
 RL J. Invest. Dermatol. 118:545-547(2002).  
 RN [12]  
 RP FUNCTION: May be a marker of basal cell differentiation in complex  
 epidermis and therefore indicative of a certain type of epithelial  
 "stem cells".  
 CC -! SUBUNIT: Heterodimer of a type I and a type II keratin. KRT17  
 associates with KRT6 isomers.  
 CC -! INTERACTION:  
 CC Q96HA0:DIPA; NBExp=1; IntAct=EBI-297873, EBI-739674;  
 CC TISSUE SPECIFICITY: Expressed in the hair follicle, nail bed,  
 mucosal stratified squamous epithelia and in basal cells of oral  
 epithelium, palmo-plantar epidermis and sweat and mammary glands.  
 CC -! INDUCTION: Induced in damaged or stressed epidermis.  
 CC -! DISEASE: Defects in KRT17 are a cause of pachyonychia congenita  
 type 2 (PC2) [MIM:167210]; also known as Jadassohn-Lewandowsky (J-  
 L) syndrome. PC2 is characterized by onychogryphosis, limited  
 plantar hyperkeratosis, multiple epidermal cysts, abnormal eyebrow  
 and body hair and by the presence of natal teeth.  
 CC -! DISEASE: Defects in KRT17 are a cause of steatocystoma multiplex  
 (SM) [MIM:184500]. SM is a disease characterized by round or oval  
 cystic tumors widely distributed on the back, anterior trunk,  
 arms, scrotum, and thighs.  
 CC -! DISEASE: KRT16 and KRT17 are coexpressed only in pathological  
 situations such as metaplasias and carcinomas of the uterine  
 cervix and in psoriasis vulgaris.  
 CC -! MISCELLANEOUS: There are two types of cytoskeletal and  
 microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
 55 and 56-70 kDa, respectively).  
 CC -! SIMILARITY: Belongs to the intermediate filament family.  
 CC -! DATABASE: NAME=Human Intermediate Filament Mutation Database;  
 WWW="http://www.interfil.org".  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 Distributed under the Creative Commons Attribution-NoDerivs License  
 -----  
 DR EMBL; Z19574; CAA79626.1; -; Genomic\_DNA.  
 DR EMBL; X62571; CAA44451.1; -; mRNA.  
 DR EMBL; S78515; AAB34565.1; -; Genomic\_DNA.  
 DR EMBL; BC000159; AAH00159.2; -; mRNA.  
 DR EMBL; BC011901; AAH11901.1; -; mRNA.  
 DR EMBL; BC056421; AAH56421.1; -; mRNA.  
 DR EMBL; BC072019; AAH72019.1; -; mRNA.  
 DR PIR; S30433; S30433.  
 DR HSSP; P08670; 1GK7.  
 DR IntAct; Q04695; -.  
 DR SWISS-2DPAGE; Q04695; HUMAN.  
 DR PHCI-2DPAGE; Q04695; -.

DR Ensembl; ENSG00000186831; Homo sapiens.  
 DR H-InvDB; HIX0013814; -.  
 DR HGNC; HGNC:6427; KRT17.  
 DR MIM; 148069; gene.  
 DR MIM; 167210; phenotype.  
 DR MIM; 184500; phenotype.  
 DR LinkHub; Q04695; -.  
 DR GO; GO:0005882; C:intermediate filament; TAS.  
 DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.  
 DR GO; GO:0008544; P:epidermis development; TAS.  
 DR InterPro; IPR001664; IF.  
 DR Pfam; PF00038; Filament; 1.  
 DR PRINTS; PR01248; TYPEKERATIN.  
 DR PROSITE; PS00226; IF; 1.  
 KW Coiled coil; Disease mutation; Intermediate filament; Keratin.  
 FT INIT MET 0 0 By similarity.  
 FT CHAIN 1 431 Keratin, type I cytoskeletal 17.  
 FT /FTID=PRO\_0000063664.  
 FT REGION 1 82 Head.  
 FT REGION 83 391 Rod.  
 FT REGION 83 119 Coil 1A.  
 FT REGION 120 137 Linker 1.  
 FT REGION 138 229 Coil 1B.  
 FT REGION 230 249 Linker 12.  
 FT REGION 250 391 Coil 2.  
 FT REGION 392 431 Tail.  
 FT VARIANT 87 87 M -> T (in PC2).  
 FT VARIANT 91 91 /FTID=VAR\_010512.  
 FT VARIANT 91 91 N -> D (in PC2).  
 FT VARIANT 91 91 /FTID=VAR\_003847.  
 FT VARIANT 91 91 N -> H (in SM).  
 FT VARIANT 91 91 /FTID=VAR\_003848.  
 FT VARIANT 91 91 N -> S (in PC2).  
 FT VARIANT 93 97 /FTID=VAR\_003849.  
 FT VARIANT 93 97 Missing (in PC2).  
 FT VARIANT 93 97 /FTID=VAR\_017069.  
 FT VARIANT 93 93 R -> C (in PC2 and SM).  
 QY 1 STMQELNSRLASVLDK 16  
 DB 85 ATMQNLDRLASVLDK 100  
 Query Match 84.4%; Score 65; DB 1; Length 431;  
 Best Local Similarity 81.2%; Pred. No. 0.023;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 RESULT 13  
 KLC17 MOUSE  
 ID KLC17\_MOUSE STANDARD; PRT; 432 AA.  
 AC Q9QWL7; Q61783;  
 DT 19-SEP-2002, integrated into UniProtKB/Swiss-Prot.  
 DT 19-SEP-2002, sequence version 2.  
 DT 07-FEB-2006, entry version 37.  
 DE Keratin, type I cytoskeletal 17 (Cytokeratin-17) (CK-17) (Keratin-17)  
 DE (K17).  
 GN Name=Krt17; Synonyms=Krt11-17;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RC STRAIN=C57BL/10J; TISSUE=Skin;  
 RX MEDLINE=99189238; PubMed=10087197; DOI=10.1006/geno.1998.5721;  
 RA Sato H., Koide T., Sagai T., Ishiguro S.I., Tamai M., Saitou N.,  
 Shiroishi T.;  
 RT "The genomic organization of type I keratin genes in mice.";  
 RL Genomics 56:303-309(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA] OF 228-432.

```

RX MEDLINE=87109202; PubMed=2433272;
RA Knapp B., Rentrop M., Schweizer J., Winter H.;
RT "Three cDNA sequences of mouse type I keratins: cellular localization
of the mRNAs in normal and hyperproliferative tissues.";
RL J. Biol. Chem. 262:938-945(1987).
RN [3]
RP DEVELOPMENTAL STAGE.
RX MEDLINE=99003303; PubMed=9786956; DOI=10.1083/jcb.143.2.469;
RA McGowan K.M., Coulombe P.A.;
RT "Onset of keratin 17 expression coincides with the definition of major
epithelial lineages during skin development.";
RL J. Cell Biol. 143:469-486(1998).
CC -!- SUBUNIT: Heterodimer of a type I and a type II keratin. KRT17
associates with KRT6 isomers (By similarity).
CC -!- DEVELOPMENTAL STAGE: Expression first occurs in a subset of
epithelial cells within the single-layered, undifferentiated
ectoderm of embryonic day 10.5 mouse fetuses. In the ensuing 48
hours, K17-expressing cells give rise to placodes, the precursors
of ectoderm-derived appendages (hair, glands, and tooth), and to
periderm.
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and
microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
55 and 56-70 kDa, respectively).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
-----
DR EMBL; AB013608; BAA34229.1; -; mRNA.
DR EMBL; M13805; AAA39394.1; -; mRNA.
DR PIR; C26135; C26135.
DR HSSP; P08670; 1GK7.
DR Ensembl; ENSMUSG00000035557; Mus musculus.
DR MGI; MGI:96691; Krt11-17.
DR LinkHub; O9QWL7; -.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF, 1.
KW Coiled coil; Intermediate filament; Keratin.
FT INIT MET 0 0
FT CHAIN 1 432
FT REGION 1 82
FT REGION 83 391
FT REGION 83 119
FT REGION 120 137
FT REGION 138 229
FT REGION 230 249
FT REGION 250 391
FT REGION 332 432
SQ SEQUENCE 432 AA; 58703AD0A02A33042 CRC64;

Query Match 84.4%; Score 65; DB 1; Length 432;
Best Local Similarity 81.2%; Pred. No. 0.023;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
Db 85 ATMQLNDRSLASYLDK 100
:|||||
RESULT 14
Q6IFU8 RAT PRELIMINARY; PRT; 433 AA.
AC Q6IFU8;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 13.
DE Type I keratin K17.
GN Name=K17;
OS Rattus norvegicus (Rat).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BN;
RX MEDLINE=23257644; PubMed=15085952; DOI=10.1078/0171-9335-00354;
RA Hesse M., Zimek A., Weber K., Magin T.M.;
RT "Comprehensive analysis of keratin gene clusters in humans and
rodents.";
RL Eur. J. Cell Biol. 83:19-26(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Thymus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Distchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Thymus;
RG NIH MGC Project;
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
EMBL/GenBank/DBJ third party annotation (TPA) entry.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
-----
DR EMBL; BK004050; DAA04484.1; -; mRNA.
DR EMBL; BC100058; AAI00059.1; -; Rattus norvegicus.
DR Ensembl; ENSRNOG0000026371; Rattus norvegicus.
DR GO; GO:0005882; C:intermediate filament; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; Filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF, 1.
KW Intermediate filament; Keratin.
SQ SEQUENCE 433 AA; 48123 MW; 7CB5C921841382D2 CRC64;

Query Match 84.4%; Score 65; DB 2; Length 433;
Best Local Similarity 81.2%; Pred. No. 0.023;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
Db 86 ATMQLNDRSLASYLDK 101
:|||||
RESULT 15
Q62253 RAT PRELIMINARY; PRT; 123 AA.
ID Q62253 RAT

```

Search completed: July 5, 2006, 19:01:24  
Job time : 165.828 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 18:28:26 ; Search time 139.029 Seconds  
(without alignments)  
52.618 Million cell updates/sec

Title: US-10-774-928A-14  
Perfect score: 77  
Sequence: 1 STMQELNRLASYLDK 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_8.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*  
9: Geneseqp2005s.\*  
10: Geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	77	100.0	16	AAO26394	Aao26394 Psoriasis
2	77	100.0	16	ADV04419	Adv04419 Leishmani
3	77	100.0	622	ADS88365	Ads88365 Human pro
4	77	100.0	622	AEA15476	Aea15476 Human pol
5	77	100.0	622	AEB13340	Aeb13340 Mutant ke
6	66	85.7	423	ADF91430	Adf91430 Mouse cyt
7	65	84.4	167	AAb43930	Aab43930 Human can
8	65	84.4	218	ADT50990	Adt50990 Cancer re
9	65	84.4	247	ADT50969	Adt50969 Cancer re
10	65	84.4	249	ADT50968	Adt50968 Cancer re
11	65	84.4	262	ADT50985	Adt50985 Cancer re
12	65	84.4	303	ADT50991	Adt50991 Cancer re
13	65	84.4	332	ADT50984	Adt50984 Cancer re
14	65	84.4	432	AAW54233	Aaw54233 Human cyt
15	65	84.4	432	ABG77168	Abg77168 Prostate
16	65	84.4	432	ADNO4112	Adn04112 Antipsori-
17	65	84.4	432	ADN9874	Adn9874 Novel hum
18	65	84.4	432	ABM81869	Abm81869 Tumour-as
19	65	84.4	432	ADV85758	Adv85758 Human pro
20	65	84.4	432	AEA15479	Aea15479 Human pol
21	65	84.4	432	AED74635	Aed74635 Human pla
22	65	84.4	433	ABM81838	Abm81838 Tumour-as
23	65	84.4	459	AAbs58761	Aabs58761 Breast an

24	65	84.4	757	4	ABG06703	Abg06703 Novel hum
25	65	84.4	887	4	ABG23684	Abg23684 Novel hum
26	65	84.4	1113	4	ABG06702	Abg06702 Novel hum
27	64	83.1	128	3	AAbs7152	Aabs7152 Human pro
28	64	83.1	142	2	AAy29566	Aay29566 Human lun
29	64	83.1	142	3	AAb44489	Aab44489 Human lun
30	64	83.1	142	4	AAe13831	Aae13831 Human lun
31	64	83.1	142	7	ADD66521	Add66521 Human lun
32	64	83.1	142	7	ADe87775	Aded87775 Human lun
33	64	83.1	166	4	ABG05346	Abg05346 Novel hum
34	64	83.1	189	8	ABO60480	Ab060480 Human gen
35	64	83.1	200	4	ADR98843	Adr98843 Lung spec
36	64	83.1	221	8	ABG23687	Abg23687 Novel hum
37	64	83.1	265	8	ADP56617	Adp56617 Human bre
38	64	83.1	281	8	ADT51001	Adt51001 Cancer re
39	64	83.1	315	8	ADP56622	Adp56622 Human bre
40	64	83.1	338	8	ADP56623	Adp56623 Human bre
41	64	83.1	388	8	ADR98903	Adr98903 Lung spec
42	64	83.1	393	8	ADT50998	Adt50998 Cancer re
43	64	83.1	400	6	ABP55382	Abp55382 Human col
44	64	83.1	400	7	ADI62966	Adi62966 Human apo
45	64	83.1	400	7	ADI62989	Adi62989 Human apo

## ALIGNMENTS

RESULT 1  
AAO26394  
ID AAO26394 standard; peptide; 16 AA.  
AC AAO26394;  
XX  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Psoriasis treating immunotherapeutic peptide, SEQ ID No 14.  
XX  
KW Antipsoriatic; immunostimulant; gene therapy; pharmaceutical composition;  
KW vaccine; immunogenic variant; immunotherapeutic agent; psoriasis;  
KW gene therapy.  
XX  
OS Leishmania sp.  
XX  
XX WO200274239-A2.  
PN  
XX  
PD 26-SEP-2002.  
XX  
PF 04-MAR-2002; 2002WO-US006496.  
XX  
PR 16-MAR-2001; 2001US-00809003.  
XX  
XX (AKIV-) AKIVA LLC.  
PI O'daly JA;  
XX  
DR WPI; 2003-018763/01.  
XX  
PT New polypeptide and immunogenic variants comprising amino acid sequences of particulate antigens, useful for the treatment and clinical remission of psoriasis.  
XX  
XX Claim 1; Page 43; 56pp; English.  
XX  
CC The invention relates to a polypeptide comprising an isolated amino acid sequence or immunogenic variants selected from any of 14 fully defined sequences of 7-16 amino acids, given in the specification. The  
CC sequences of 7-16 amino acids, given in the specification. The  
CC immunotherapeutic agents and a pharmaceutical compositions comprising polynucleotides and vectors of the invention are useful for the treatment and clinical remission of psoriasis. The isolated nucleic acids are  
CC useful as probes. The sequences of the invention can be used in the treatment of disorders by gene therapy. This sequence represents one of the 14 immunotherapeutic peptides of the invention  
XX

```
SQ      Sequence 16 AA;
Query Match      100.0%; Score 77; DB 6; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 STMQELNSRLASYLDK 16
      |||||
DB      1 STMQELNSRLASYLDK 16
      |||||

RESULT 2
ADV04419
ID      ADV04419 standard; peptide; 16 AA.
XX
AC      ADV04419;
XX
DT      24-FEB-2005 (first entry)
XX
DE      Leishmania immunotherapeutic peptide SEQ ID NO:14.
XX
KW      T-lymphocyte; immunogenicity; antigen; antipsoriatic; immunostimulant;
KW      psoriasis.
XX
OS      Leishmania sp.
XX
PN      US2004241168-A1.
XX
PD      02-DEC-2004.
XX
PF      09-FEB-2004; 2004US-00774928.
XX
PR      16-MAR-2001; 2001US-00809003.
XX
PR      17-OCT-2003; 2003US-00687892.
XX
PA      (ODAL/) ODALY J A.
PI      Odaly JA;
XX
WPI;    2005-011563/01.
XX
PT      Inhibiting selectively T-cell rolling in human, by administering compound
PT      interfering with cutaneous lymphocyte antigen-E selectin interaction and
PT      leukocyte function associated antigen-1 interaction.
XX
PS      Claim 18; SEQ ID NO 14; 21pp; English.
XX
CC      The invention relates to a novel method for selectively inhibiting T-cell
CC      rolling in a human host, comprising administering a compound that
CC      selectively interferes with the cutaneous lymphocyte antigen (CLA)-E
CC      selectin interaction and leukocyte function associated antigen (LFA)-
CC      1/intercellular adhesion molecule (ICAM) and very late antigen
CC      (VLA)/vascular cell adhesion molecule (VCAM) interactions. The compound
CC      includes an immunotherapeutic agent, which comprises a purified protein
CC      extract that is isolated by diethylaminoethyl Sephadex chromatography of
CC      a Nonidet P-40 insoluble particulate antigen fraction derived from
CC      isolated killed cells of anastigotes from one or more species of the
CC      Leishmania genus, where the particulate antigen fraction is solubilized
CC      with 8M urea and 0.025 M Tris(hydroxymethyl)aminomethane pH 8.3 applied
CC      to diethylaminoethyl Sephadex and is eluted with a solution comprising
CC      0.1 M sodium chloride, 8 M urea and 0.025 M
CC      Tris(hydroxymethyl)aminomethane pH 8.3, and the purified protein extract
CC      includes polypeptides having apparent molecular weights of 73, 80 and 82
CC      kDa, after total reduction and alkylation. The species is L. amazonensis,
CC      L. venezuelensis, L. brasiliensis and/or L. chagasi. A compound of the
CC      invention has antipsoriatic and immunostimulant activity, and inhibits T-
CC      cell rolling by interfering with CLA-E selectin interaction and LFA-
CC      1/ICAM and VLA/VCAM interactions. The method is useful for selectively
CC      inhibiting T-cell rolling in a human host, and for treating psoriasis.
CC      The present sequence represents a Leishmania peptide used in the
CC      invention.
XX
SQ      Sequence 16 AA;
Query Match      100.0%; Score 77; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 STMQELNSRLASYLDK 16
      |||||
DB      1 STMQELNSRLASYLDK 16
      |||||

RESULT 3
ADS88365
ID      ADS88365 standard; protein; 622 AA.
XX
AC      ADS88365;
XX
DT      18-NOV-2004 (first entry)
XX
DE      Human protein of a TNF-alpha signalling pathway protein complex Seq 220.
XX
KW      protein complex; tumour necrosis factor-alpha signalling pathway;
KW      TNF-alpha; chronic inflammatory disease; rheumatoid arthritis;
KW      inflammatory bowel disease; infectious disease; septic shock;
KW      bacterial infection; neurological disease; stroke-induced inflammation;
KW      neurodegenerative disease; cancer; antiinflammatory; antiarthritic;
KW      antirheumatic; cytostatic; antibacterial; gene therapy; human.
XX
OS      Homo sapiens.
XX
PN      WO2004035783-A2.
XX
PD      29-APR-2004.
XX
PF      24-SEP-2003; 2003WO-EP050655.
XX
PR      26-SEP-2002; 2002EP-00021809.
XX
PR      10-FEB-2003; 2003EP-00100274.
XX
PA      (CELL-) CELLZOME AG.
XX
PI      Bouwmeester T, Huhse B, Bauch A, Ruffner H, Bauer A, Kuester B;
PI      Superti-Furga G, Kruse U;
XX
WPI;    2004-348460/32.
XX
PT      New protein complex comprising at least one first and second protein of
PT      the Tumor Necrosis Factor-alpha(TNF-alpha)-signaling pathway, useful for
PT      diagnosing or treating inflammation, neurological diseases, infectious
PT      diseases or cancer.
XX
PS      Example; SEQ ID NO 220; 1980pp; English.
XX
CC      This invention relates to novel protein complexes of the tumour necrosis
CC      factor-alpha (TNF-alpha) signalling pathway. Specifically, it refers to
CC      methods for preparing these complexes comprising at least two component
CC      proteins, as well as screening methods to identify modulators of the
CC      pathway, which include antibodies, agonists and antagonists thereof. The
CC      present invention describes a protein complex and kit that are useful for
CC      diagnosing, prognosing or treating chronic inflammatory diseases such as
CC      rheumatoid arthritis and inflammatory bowel disease; infectious diseases
CC      such as septic shock and bacterial infections; neurological diseases such
CC      as stroke-induced inflammation in neurons; neurodegenerative diseases and
CC      cancer. Accordingly, these complexes can be used for the development of
CC      pharmaceutical compositions that exhibit antiinflammatory, antiarthritic,
CC      antirheumatic, cytostatic and antibacterial activities and can be used
CC      for gene therapy purposes. In particular, the invention further provides
CC      siRNA-oligonucleotides useful for inhibiting protein expression for in
CC      vitro or cell culture assays. This polypeptide is a human protein that
CC      can be used in combination with other proteins provided in the
CC      specification to form novel complexes of the TNF-alpha signalling pathway
CC      of the invention.
XX
SQ      Sequence 622 AA;
```

Query Match 100.0%; Score 77; DB 8; Length 622;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16  
|||||  
Db 154 STMQELNSRLASYLDK 169

RESULT 4  
AEA15476  
ID AEA15476 standard; protein; 622 AA.  
XX  
AC AEA15476;  
XX  
DT 28-JUL-2005 (first entry)  
XX  
DE Human polypeptide #95.  
XX  
KW Diagnosis; prognosis; cancer; breast tumor; ovary tumor; stomach tumor;  
KW colon tumor; esophagus tumor; bladder tumor; non-small-cell lung cancer;  
KW cytostatic; neoplasm.  
XX  
OS Homo sapiens.  
XX  
PN WO2005047534-A2.  
XX  
PD 26-MAY-2005.  
XX  
PF 15-OCT-2004; 2004WO-EP011599.  
XX  
PR 28-OCT-2003; 2003EP-00024565.  
XX  
PA (PAB ) BAYER HEALTHCARE AG.  
XX  
PI Wirtz R, Munnes M;  
XX  
DR WPI; 2005-372393/38.  
XX  
DR N-PSDB; AEA15398.

XX  
PT Predicting a response to cancer treatment by detecting at least 2  
PT markers, which are genes or genomic nucleic acid sequences that are  
PT located on one chromosomal region, which is altered in malignant  
PT neoplasia.

PS Claim 7; SEQ ID NO 438; 464pp; English.

XX  
CC The invention relates to a method of predicting response to cancer  
CC treatment comprising detection of at least 2 markers, where the markers  
CC are genes and fragments or genomic nucleic acid sequences that are  
CC located on one chromosomal region, which is altered in malignant  
CC neoplasia. The invention also relates to a method for the prediction,  
CC diagnosis or prognosis of malignant neoplasia, methods for detecting  
CC deregulations in malignant neoplasia and breast cancer, a method of  
CC determining the phenotype of a cell or tissue, a method for identifying  
CC genomic regions which are altered on the chromosomal level and encode  
CC genes that are linked by function and are differentially expressed in  
CC malignant neoplasia and breast cancer, methods of screening for agents  
CC which regulate the activity of a polypeptide or a polynucleotide and  
CC antibodies that specifically bind to a full length or partial  
CC polypeptide. The method is useful for predicting response to cancer  
CC treatment. The methods and compositions are useful for predicting,  
CC diagnosing, prognosing, preventing or treating malignant neoplasia  
CC including breast cancer, ovarian cancer, gastric cancer, colon cancer,  
CC esophageal cancer, mesenchymal cancer, bladder cancer or non-small-cell  
CC lung cancer. This sequence represents a human polypeptide used in the  
CC scope of the invention.

SQ Sequence 622 AA;

Query Match 100.0%; Score 77; DB 9; Length 622;  
Best Local Similarity 100.0%; Pred. No. 0.00011;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 STMQELNSRLASYLDK 16  
|||||  
Db 154 STMQELNSRLASYLDK 169

RESULT 5  
AEB13340  
ID AEB13340 standard; protein; 622 AA.  
XX  
AC AEB13340;  
XX  
DT 08-SEP-2005 (first entry)  
XX  
DE Mutant keratin-9 SEQ ID NO 5.  
XX  
KW antigen; tumor-associated antigen; diagnostic; therapeutic; tumor marker;  
KW squamous cell carcinoma;  
KW autoantibody-mediated identification of antigens; AMIDA; cytostatic;  
KW neoplasm; cancer; skin tumor; ear, nose, throat disease; mutant;  
KW keratin-9.  
XX  
OS Homo sapiens.  
OS Synthetic.

PN WO2005061537-A2.

XX  
PD 07-JUL-2005.

XX  
PF 22-DEC-2004; 2004WO-EP014634.

XX  
PR 22-DEC-2003; 2003DE-01060456.

XX  
PA (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEI.  
PA (VAEC-) VAECGENE BIOTECH GMBH.

XX  
PI Gires O, Zeidler R, Muenz M, Schaffrik M;

XX  
DR WPI; 2005-479443/48.

XX  
PT Antigen for skin tumors in the nose and throat, uses at least one protein  
PT from a selected sequence or their variants with a hybridized isolated  
PT nucleic acid encoding the transcribed product.

PS Claim 1; SEQ ID NO 5; 54pp; German.

XX  
CC This invention describes the detection of novel antigens associated with  
CC squamous cell carcinomas using autoantibody-mediated identification of  
CC antigens (AMIDA). The invention also describes methods for producing  
CC antigen-presenting cells and T cells that are specific for the antigens  
CC described. The antigens can be used in a diagnostic formulation or as a  
CC pharmaceutical product for therapy or as a vaccine, for diagnosis and  
CC therapy of squamous cell tumors and skin tumors associated with the ear,  
CC nose and throat (ENT). The tumor antigen is an improved diagnostic marker  
CC for tumor identification, and is also an effective therapeutic agent.  
CC This sequence represents the tumor-associated antigen corresponding to a  
CC mutant keratin-9.

SQ Sequence 622 AA;

Query Match 100.0%; Score 77; DB 9; Length 622;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16  
|||||  
Db 154 STMQELNSRLASYLDK 169

RESULT 6  
ADF91430  
ID ADF91430 standard; protein; 423 AA.

```

XX AC ADF91430;
XX DT 26-FEB-2004 (first entry)
XX DE Mouse cytokeatin 18 #SEQ ID 2.
XX KW Antiasthmatic; anti-allergic; anti-inflammatory; gene therapy;
XX KW bronchial asthma; chronic rhinitis; cytokeatin 18.
XX OS Mus sp.
XX PN WO2003098211-A1.
XX PD 27-NOV-2003.
XX PF 12-MAY-2003; 2003WO-KR000933.
XX PR 15-MAY-2002; 2002KR-00026765.
XX PA (NAHM/) NAHM D.
XX PA (JEON/) JEON S.
XX PI Nahm D, Jeon S;
XX WPI; 2004-022911/02.
XX DR
XX XX Diagnosing bronchial asthma and chronic rhinitis comprises detecting
PT autoantibodies to cytokeatin 18 protein in the bodily fluid of a human
PT subject.
XX PS Claim 18; SEQ ID NO 2; 42pp; English.
XX CC The invention relates to a method for diagnosing bronchial asthma and
CC chronic rhinitis comprising detecting autoantibodies to cytokeatin 18
CC protein in the bodily fluid of a human subject. The method or cytokeatin
CC 18 protein is useful in diagnosing, classifying bronchial asthma and
CC chronic rhinitis or for formulating pharmaceutical formulations for
CC protecting or treating patients with or non-allergic patients with
CC bronchial asthma and chronic rhinitis or patients with bronchial asthma
CC and chronic rhinitis associated with autoantibodies to cytokeatin 18.
CC Cytokeatin 18 protein is used as drug target in the production of drugs
CC for treating bronchial asthma and chronic rhinitis. The current sequence
CC represents mouse cytokeatin 18.
XX SQ Sequence 423 AA;
    Query Match      85.7%; Score 66; DB 8; Length 423;
    Best Local Similarity 86.7%; Pred. No. 0.0067;
    Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TMQELNSRLASVLDK 16
    |||:|||||
DB 75 TMQDLNDRLASVLDK 89

RESULT 7
AAB43930
ID AAB43930 standard; protein; 167 AA.
XX AC AAB43930;
XX DT 08-FEB-2001 (first entry)
XX DE Human cancer associated protein sequence SEQ ID NO:1375.
XX KW Human; cancer associated gene; cancer antigen; detection; cancer;
XX KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
XX KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
XX KW anti-inflammatory; antithyroid; antiallergic; antibacterial; cardiant;
XX KW dermatological; neuroprotective; thrombolytic; coagulant; neutropic;
XX KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
XX KW immune disorder; haematopoietic cell disorder; autoimmune disorder;

```

```

KW KW allergic reaction; graft versus host disease; organ rejection;
KW KW haemostatic; thrombolytic; cardiovascular disorder; infection;
XX OS Homo sapiens.
XX PN WO200055350-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005882.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX WPI; 2000-587533/55.
XX N-PSDB; AAC78139.
XX PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer.
XX PS Claim 11; Page 2040-2041; 2352pp; English.
XX CC AAC7807 to AAC78448 encode the human cancer associated proteins given in
CC AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerable; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC anti-inflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC neutropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention
XX SQ Sequence 167 AA;
    Query Match      84.4%; Score 65; DB 3; Length 167;
    Best Local Similarity 86.7%; Pred. No. 0.0034;
    Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TMQELNSRLASVLDK 16
    |||:|||||
DB 118 TMQDLNDRLASVLDK 132

RESULT 8
ADT50990
ID ADT50990 standard; protein; 218 AA.
XX AC ADT50990;
XX DT 13-JAN-2005 (first entry)
XX DE Cancer related protein sequence #153.
XX KW Cytostatic; gene therapy; vaccine; diagnosis; breast; colon; lung;
KW ovarian; prostate; cancer.
XX OS Homo sapiens.

```

PT New cancer specific nucleic acid (CasNA) molecules, useful for  
PT diagnosing, monitoring the presence of, or treating a patient with

```

QY 1 STMQELNSRLASYLDK 16
Db 86 ATMQNLDRLASYLDK 101

RESULT 11
ADT50985
ID ADT50985 standard; protein; 262 AA.
XX
AC ADT50985;
XX
DT 13-JAN-2005 (first entry)
XX
DE Cancer related protein sequence #148.
XX
KW cytostatic; gene therapy; vaccine; diagnosis; breast; colon; lung;
KW ovarian; prostate; cancer.
OS Homo sapiens.
XX
PN WO2004092338-A2.
XX
PD 28-OCT-2004.
XX
PF 12-APR-2004; 2004WO-US011104.
XX
PR 11-APR-2003; 2003US-0462399P.
XX
PT 01-JUL-2003; 2003US-0484333P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Macina RA, Turner LR, Sun Y, Tam A;
XX
DR WPI; 2004-766851/75.
XX
PT New cancer specific nucleic acid (CasNA) molecules, useful for
PT diagnosing, monitoring the presence of, or treating a patient with
PT breast, colon, lung, ovarian, or prostate cancer.
XX
OS Homo sapiens.
XX
PN WO2004092338-A2.
XX
PD 28-OCT-2004.
XX
PF 12-APR-2004; 2004WO-US011104.
XX
PR 11-APR-2003; 2003US-0462399P.
XX
PT 01-JUL-2003; 2003US-0484333P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Macina RA, Turner LR, Sun Y, Tam A;
XX
DR WPI; 2004-766851/75.
XX
PT New cancer specific nucleic acid (CasNA) molecules, useful for
PT diagnosing, monitoring the presence of, or treating a patient with
PT breast, colon, lung, ovarian, or prostate cancer.
XX
PS Claim 12; SEQ ID NO 289; 891pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I)
CC selectively hybridizing to, or comprising at least 95% sequence identity
CC to, any of the 362 nucleotide sequences fully defined in the
CC specification. The nucleic acid molecules and polypeptides are useful for
CC diagnosing, monitoring the presence of, or treating a patient with
CC breast, colon, lung, ovarian, or prostate cancer. This sequence
CC corresponds to a protein of the invention.
XX
SQ Sequence 262 AA;

Query Match 84.4%; Score 65; DB 8; Length 262;
Best Local Similarity 81.2%; Pred. No. 0.0057;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
Db 114 ATMQNLDRLASYLDK 129

RESULT 12
ADT50991
ID ADT50991 standard; protein; 303 AA.
XX
AC ADT50991;
XX
DT 13-JAN-2005 (first entry)
XX
DE Cancer related protein sequence #154.
XX
KW cytostatic; gene therapy; vaccine; diagnosis; breast; colon; lung;
KW ovarian; prostate; cancer.
XX

```

```

OS Homo sapiens.
XX
PN WO2004092338-A2.
XX
PD 28-OCT-2004.
XX
PF 12-APR-2004; 2004WO-US011104.
XX
PR 11-APR-2003; 2003US-0462399P.
XX
PT 01-JUL-2003; 2003US-0484333P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Macina RA, Turner LR, Sun Y, Tam A;
XX
DR WPI; 2004-766851/75.
XX
PT New cancer specific nucleic acid (CasNA) molecules, useful for
PT diagnosing, monitoring the presence of, or treating a patient with
PT breast, colon, lung, ovarian, or prostate cancer.
XX
PS Claim 12; SEQ ID NO 295; 891pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I)
CC selectively hybridizing to, or comprising at least 95% sequence identity
CC to, any of the 362 nucleotide sequences fully defined in the
CC specification. The nucleic acid molecules and polypeptides are useful for
CC diagnosing, monitoring the presence of, or treating a patient with
CC breast, colon, lung, ovarian, or prostate cancer. This sequence
CC corresponds to a protein of the invention.
XX
SQ Sequence 303 AA;

Query Match 84.4%; Score 65; DB 8; Length 303;
Best Local Similarity 81.2%; Pred. No. 0.0068;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
Db 133 ATMQNLDRLASYLDK 148

RESULT 13
ADT50984
ID ADT50984 standard; protein; 332 AA.
XX
AC ADT50984;
XX
DT 13-JAN-2005 (first entry)
XX
DE Cancer related protein sequence #147.
XX
KW cytostatic; gene therapy; vaccine; diagnosis; breast; colon; lung;
KW ovarian; prostate; cancer.
XX
OS Homo sapiens.
XX
PN WO2004092338-A2.
XX
PD 28-OCT-2004.
XX
PF 12-APR-2004; 2004WO-US011104.
XX
PR 11-APR-2003; 2003US-0462399P.
XX
PT 01-JUL-2003; 2003US-0484333P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Macina RA, Turner LR, Sun Y, Tam A;
XX
DR WPI; 2004-766851/75.
XX
PT New cancer specific nucleic acid (CasNA) molecules, useful for

```

PT diagnosing, monitoring the presence of, or treating a patient with  
 PT breast, colon, lung, ovarian, or prostate cancer.  
 PS Claim 12; SEQ ID NO 288; 891pp; English.  
 XS  
 CC The invention relates to an isolated nucleic acid molecule (I)  
 CC selectively hybridizing to, or comprising at least 95% sequence identity  
 CC to, any of the 362 nucleotide sequences fully defined in the  
 CC specification. The nucleic acid molecules and polypeptides are useful for  
 CC diagnosing, monitoring the presence of, or treating a patient with  
 CC breast, colon, lung, ovarian, or prostate cancer. This sequence  
 CC corresponds to a protein of the invention.  
 XX  
 SQ Sequence 332 AA;  
 Query Match 84.4%; Score 65; DB 8; Length 332;  
 Best Local Similarity 81.2%; Pred. No. 0.0076; 2; Indels 0; Gaps 0;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 STMQELNSRLASYLDK 16  
 Db 117 ATMQLNDRSLASYLDK 132  
 RESULT 14  
 AAW54233  
 ID AAW54233 standard; peptide; 432 AA.  
 AC AAW54233;  
 DT 10-AUG-1998 (first entry)  
 DE Human cytokeratin 17 protein.  
 KW Cervical cancer-associated protein; Cvc; tryptic peptide; human;  
 KW detection; treatment; cytokeratin 17; non-chromatin protein.  
 XX Homo sapiens.  
 OS  
 PN WO9809170-A2.  
 XX  
 PD 05-MAR-1998.  
 XX  
 PF 19-AUG-1997; 97WO-US014526.  
 XX  
 PR 30-AUG-1996; 96US-00705660.  
 PA (MATR-) MATRITECH INC.  
 XX  
 PI Keesee SK, Obar R, Wu Y;  
 XX  
 DR WPI; 1998-230271/20.  
 XX  
 PT Detection and therapy of cervical cancer - using specific cervical cancer  
 PT -associated proteins as targets for treatment or as indicators for  
 PT detection.  
 PS Claim 10; Page 52-53; 79pp; English.  
 XX  
 CC This protein is the human cytokeratin 17 which is used to obtain tryptic  
 CC peptides which are used in a method for detecting cervical cancer. The  
 CC method involves detecting the presence of a cervical cancer-associated  
 CC protein (Cvc) in a tissue or body fluid sample. The Cvc is characterised  
 CC as having a molecular weight of 44900-69400 Daltons as determined by  
 CC sodium dodecyl-sulphate (SDS)-PAGE techniques and an isoelectric point  
 CC (pI) of 5.1-6.6 as determined by standard isoelectric focusing  
 CC techniques. The protein is further characterised as being a non-chromatin  
 CC protein which is detectable at a higher level in a human cervical cancer  
 CC cell than in a normal human cervical cell, as determined by 2D-gel  
 CC electrophoresis. The methods can be used for the early and rapid  
 CC detection of cervical cancer, for treating cervical cancers and for  
 CC monitoring the efficacy of such treatment

SQ Sequence 432 AA;  
 Query Match 84.4%; Score 65; DB 2; Length 432;  
 Best Local Similarity 81.2%; Pred. No. 0.01; 2; Indels 0; Gaps 0;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 STMQELNSRLASYLDK 16  
 Db 86 ATMQLNDRSLASYLDK 101  
 RESULT 15  
 ABG77168  
 ID ABG77168 standard; protein; 432 AA.  
 XX  
 AC ABG77168;  
 DT 05-NOV-2002 (first entry)  
 DE Prostate adenocarcinoma associated protein #3.  
 KW Prostate cancer; differential expression; cancer diagnosis;  
 KW cancer treatment; cancer monitoring; prostate adenocarcinoma.  
 XX Homo sapiens.  
 OS  
 PN US2002119463-A1.  
 XX  
 PD 29-AUG-2002.  
 XX  
 PF 30-JUL-2001; 2001US-00919172.  
 XX  
 PR 28-JUL-2000; 2000US-0222469P.  
 XX  
 PA (FARI/) FARIS M.  
 PA (TURN/) TURNER C M.  
 XX  
 PI Paris M, Turner CM;  
 XX  
 DR WPI; 2002-608155/65.  
 DR N-PSDB; ABS62731.  
 XX  
 PT New composition, useful for treating and diagnosing prostate cancer,  
 PT comprises human cDNAs that are differentially expressed in prostate  
 PT cancer.  
 PS Example; Page 29-30; 147pp; English.  
 XX  
 CC The invention describes a composition of a number of human cDNAs that are  
 CC differentially expressed in prostate cancer. The composition of the  
 CC invention useful for a high-throughput method for detecting differential  
 CC expression of cDNAs in a nucleic acid containing sample comprising  
 CC amplifying the nucleic acids of the sample, hybridising the composition  
 CC with nucleic acids of the sample, detecting the hybridisation complexes  
 CC and comparing the complexes with those of a standard, where differences  
 CC indicates differential expression. The sample is from a subject with  
 CC prostate cancer and comparison with a standard defines an early, mid or  
 CC late stage of the disease. The composition and or protein encoded by the  
 CC cDNA's are useful for a high-throughput method to screen molecules or  
 CC compounds to identify a ligand which specifically binds a cDNA or protein  
 CC comprising between cDNA/protein and molecule or compound. The molecules and  
 CC compounds are selected from DNA, RNA, peptide nucleic acid molecules,  
 CC mimetics, peptides, proteins, (ant)agonists, antibodies, immunoglobulins,  
 CC inhibitors, drug compounds, pharmaceutical agents, transcription factors,  
 CC repressors, and regulatory proteins. The composition is useful for  
 CC diagnosing, treating or monitoring the progression or treatment of  
 CC prostate cancer. The antibodies are also useful for the diagnosis of  
 CC disease. This sequence represents a prostate adenocarcinoma protein  
 XX  
 SQ Sequence 432 AA;  
 Query Match 84.4%; Score 65; DB 5; Length 432;

Best Local Similarity 81.2%; Pred. No. 0.01;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STMQELNSRLASYLDK 16  
:|||||  
Db 86 ATMQLNDRLASYLDK 101

Search completed: July 5, 2006, 18:43:45  
Job time : 141.029 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:38:52 ; Search time 117.126 Seconds  
(without alignments)  
63.277 Million cell updates/sec

Title: US-10-774-928a-14  
Perfect score: 77  
Sequence: 1 STMQELNSRLASYLDK 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:  
1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pcp.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	100.0	16	5	US-10-774-928-14
2	66	85.7	423	4	US-10-080-334-269
3	66	85.7	423	5	US-10-514-932-2
4	65	84.4	167	3	US-09-925-301-1375
5	65	84.4	432	3	US-09-919-172-9
6	65	84.4	432	3	US-09-315-355-18
7	65	84.4	432	5	US-10-752-986-9
8	65	84.4	432	5	US-10-848-572-18
9	65	84.4	432	5	US-10-821-234-1463
10	65	84.4	459	3	US-09-925-298-469
11	65	84.4	459	3	US-10-102-806-469
12	65	84.4	757	5	US-10-450-763-37062
13	65	84.4	887	5	US-10-450-763-54043
14	65	84.4	1113	5	US-10-450-763-37061
15	64	83.1	128	3	US-09-925-300-1730
16	64	83.1	142	3	US-09-738-973-213
17	64	83.1	142	3	US-09-854-133-213
18	64	83.1	142	4	US-10-144-649A-213
19	64	83.1	166	5	US-10-450-763-35705
20	64	83.1	189	4	US-10-029-386-34114
21	64	83.1	221	5	US-10-450-763-54046
22	64	83.1	400	3	US-09-922-217-1115
23	64	83.1	400	3	US-10-025-380-1115
24	64	83.1	400	5	US-10-734-564-118
25	64	83.1	400	5	US-10-733-969A-33
26	64	83.1	400	6	US-11-108-172-1115
27	64	83.1	433	5	US-10-450-763-35709

28 64 83.1 452 5 US-10-631-467-1588 Sequence 1588, Ap  
29 64 83.1 456 3 US-09-919-172-31 Sequence 31, Appl  
30 64 83.1 456 4 US-10-363-616-310 Sequence 310, App  
31 64 83.1 456 5 US-10-752-986-31 Sequence 31, Appl  
32 64 83.1 456 5 US-10-631-467-873 Sequence 873, App  
33 64 83.1 470 5 US-10-631-467-657 Sequence 657, App  
34 64 83.1 472 6 US-11-037-713-35 Sequence 35, Appl  
35 64 83.1 494 4 US-10-433-696-51 Sequence 51, Appl  
36 64 83.1 558 5 US-10-450-763-54047 Sequence 54047, A  
37 64 83.1 593 4 US-10-435-696-50 Sequence 50, Appl  
38 64 83.1 618 3 US-09-925-300-1381 Sequence 1381, Ap  
39 61 79.2 53 5 US-10-883-020-4 Sequence 4, Appl  
40 61 79.2 81 5 US-10-450-763-57390 Sequence 57390, A  
41 61 79.2 99 4 US-10-029-386-34091 Sequence 34091, A  
42 61 79.2 125 5 US-10-450-763-49053 Sequence 49053, A  
43 61 79.2 140 5 US-10-450-763-48990 Sequence 48990, A  
44 61 79.2 157 3 US-09-864-761-44299 Sequence 44299, A  
45 61 79.2 177 5 US-10-450-763-31405 Sequence 31405, A

## ALIGNMENTS

RESULT 1  
US-10-774-928-14  
; Sequence 14, Application US/10774928  
; Publication No. US20040241168A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis  
; FILE REFERENCE: 302178  
; CURRENT APPLICATION NUMBER: US/10/774,928  
; CURRENT FILING DATE: 2004-02-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 14  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Leishmania  
US-10-774-928-14

Query Match 100.0%; Score 77; DB 5; Length 16;  
Best Local Similarity 100.0%; Pred. No. 5.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16  
DB 1 STMQELNSRLASYLDK 16

RESULT 2  
US-10-080-334-269  
; Sequence 269, Application US/10080334  
; Publication No. US20040002584A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol E. A.  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Li, Li  
; APPLICANT: Shenoy, Suresh G  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Casman, Vladimir Y  
; APPLICANT: Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Padigaru, Muralidhara

```
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-08-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-08-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 269
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-080-334-269

Query Match      85.7%; Score 66; DB 4; Length 423;
Best Local Similarity 86.7%; Pred. No. 0.014;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2  TMQELNSRLASYLDK 16
Db      75  TMQDLNRLASYLDK 89

RESULT 3
US-10-514-932-2
; Sequence 2, Application US/10514932
; Publication No. US20050208583A1
; GENERAL INFORMATION:
; APPLICANT: NAHM, DONG HO
; APPLICANT: JEON, SOOK YEONG
; TITLE OF INVENTION: DETECTION OF AUTOANTIBODIES TO CYTOKERATIN 18 PROTEIN IN PATIENTS
; TITLE OF INVENTION: WITH BRONCHIAL ASTHMA AND CHRONIC RHINITIS, AND ITS APPLICATIONS
; TITLE OF INVENTION: INCLUDING A KIT FOR DIAGNOSING BRONCHIAL ASTHMA AND CHRONIC
; TITLE OF INVENTION: RHINITIS COMPRISING MAMMALIAN CYTOKERATIN 18 PROTEIN
; FILE REFERENCE: NPP03-0002
; CURRENT APPLICATION NUMBER: US/10/514,932
; CURRENT FILING DATE: 2004-11-12
```

```
; PRIOR APPLICATION NUMBER: KR1020020026765
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: MOUSE
US-10-514-932-2

Query Match      85.7%; Score 66; DB 5; Length 423;
Best Local Similarity 86.7%; Pred. No. 0.014;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2  TMQELNSRLASYLDK 16
Db      75  TMQDLNRLASYLDK 89

RESULT 4
US-09-925-301-1375
; Sequence 1375, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1634
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1375
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (161)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (163)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1375

Query Match      84.4%; Score 65; DB 3; Length 167;
Best Local Similarity 86.7%; Pred. No. 0.0074;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2  TMQELNSRLASYLDK 16
Db      118  TMQKLNRLASYLDK 132

RESULT 5
US-09-919-172-9
; Sequence 9, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
```



```
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1463
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1463

Query Match      84.4%; Score 65; DB 5; Length 432;
Best Local Similarity 81.2%; Pred. No. 0.02;
Matches 13; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
DB 86 ATMQLNDRSLASYLDK 101

RESULT 10
US-09-925-298-469
; Sequence 469, Application US/09925298
; Publication No. US20020039764A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103
; CURRENT APPLICATION NUMBER: US/09/925,298
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 469
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-298-469

Query Match      84.4%; Score 65; DB 3; Length 459;
Best Local Similarity 81.2%; Pred. No. 0.022;
Matches 13; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
DB 113 ATMQLNDRSLASYLDK 128

RESULT 11
US-10-102-806-469
; Sequence 469, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
```

```
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 469
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-806-469

Query Match      84.4%; Score 65; DB 4; Length 459;
Best Local Similarity 81.2%; Pred. No. 0.022;
Matches 13; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
DB 113 ATMQLNDRSLASYLDK 128

RESULT 12
US-10-450-763-37062
; Sequence 37062, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 37062
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: DOMAIN
; LOCATION: (577)..(602)
; OTHER INFORMATION: Vertebrate galactoside-binding lectin proteins domain
; OTHER INFORMATION: identified by eMATRIX, accession number BL00309C, p-value=7.429e-
; OTHER INFORMATION: 22, raw score of 18.65
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (84)..(388)
; OTHER INFORMATION: Intermediate filament proteins domain identified by PFam,
; OTHER INFORMATION: accession name filament, E-value=2.4e-123, PFam score of 423.2
US-10-450-763-37062

Query Match      84.4%; Score 65; DB 5; Length 757;
Best Local Similarity 81.2%; Pred. No. 0.037;
Matches 13; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
DB 87 ATMQLNDRSLASYLDK 102

RESULT 13
US-10-450-763-54043
; Sequence 54043, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
```

;; PRIOR APPLICATION NUMBER: PCT/US01/08631  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 09/540,217  
;; PRIOR FILING DATE: 2000-03-31  
;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO 54043  
;; LENGTH: 887  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (547)..(567)  
;; OTHER INFORMATION: kw ALLERGEN POLLEN CIM1 HOL-LI domain identified by eMATRIX,  
;; OTHER INFORMATION: accession number DM01724, p-value=8.091e-11, raw score of 8.14  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (84)..(857)  
;; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
;; OTHER INFORMATION: accession name filament, E-value=2e-247, Pfam score of 835.4  
US-10-450-763-54043

Query Match 84.4%; Score 65; DB 5; Length 887;  
Best Local Similarity 81.2%; Pred. No. 0.044;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 STMQELNSRLASYLDK 16  
:|||||  
DB 87 ATMQNLNDRSLASYLDK 102

## RESULT 14

US-10-450-763-37061  
;; Sequence 37061, Application US/10450763  
;; Publication No. US20050196754A1  
;; GENERAL INFORMATION:

;; APPLICANT: Hyseq, Inc  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
;; FILE REFERENCE: 790CIP3/US  
;; CURRENT APPLICATION NUMBER: US/10/450,763  
;; CURRENT FILING DATE: 2003-06-11  
;; PRIOR APPLICATION NUMBER: PCT/US01/08631  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 09/540,217  
;; PRIOR FILING DATE: 2000-03-31  
;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO 37061  
;; LENGTH: 1113  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (1004)..(1024)  
;; OTHER INFORMATION: kw ALLERGEN POLLEN CIM1 HOL-LI domain identified by eMATRIX,  
;; OTHER INFORMATION: accession number DM01724, p-value=3.045e-11, raw score of 8.14  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (84)..(1104)  
;; OTHER INFORMATION: Intermediate filament proteins domain identified by Pfam,  
;; OTHER INFORMATION: accession name filament, E-value=3.3e-277, Pfam score of 934.3  
US-10-450-763-37061

Query Match 84.4%; Score 65; DB 5; Length 1113;  
Best Local Similarity 81.2%; Pred. No. 0.057;  
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16  
:|||||

Db 87 ATMQNLNDRSLASYLDK 102

## RESULT 15

US-09-925-300-1730  
;; Sequence 1730, Application US/09925300  
;; Patent No. US20020151681A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Craig Rosen,  
;; APPLICANT: Steve Ruben  
;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
;; FILE REFERENCE: PA101  
;; CURRENT APPLICATION NUMBER: US/09/925,300  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCT/US00/05988  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1890  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1730  
;; LENGTH: 128  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-925-300-1730

Query Match 83.1%; Score 64; DB 3; Length 128;  
Best Local Similarity 86.7%; Pred. No. 0.0082;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMQELNSRLASYLDK 16  
:|||||

Db 100 TMQNLNDRSLASYLDK 114  
:|||||

Search completed: July 5, 2006, 20:51:56  
Job time: 118.126 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 20:39:32 ; Search time 9.16505 Seconds  
(without alignments)  
46.842 Million cell updates/sec

Title: US-10-774-928A-14

Perfect score: 77

Sequence: 1 STWQELNSRLASYLDK 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112942 seqs, 26832045 residues

Total number of hits satisfying chosen parameters: 112942

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New.\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pap.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	64	83.1	400	7	US-11-105-233-182
2	61	79.2	430	7	US-11-105-233-181
3	58	75.3	422	7	US-11-105-233-176
4	56	72.7	259	7	US-11-293-697-4615
5	55	71.4	424	7	US-11-105-233-183
6	52	67.5	466	7	US-11-289-102-289
7	52	67.5	551	7	US-11-270-040-4
8	49	63.6	586	6	US-10-505-928-314
9	48	62.3	450	7	US-11-293-697-4090
10	43	55.8	794	6	US-10-449-902-50320
11	43	55.8	969	6	US-10-449-902-47888
12	41	53.2	535	7	US-11-293-697-4336
13	38	49.4	328	7	US-11-293-697-3859
14	38	49.4	422	6	US-10-449-902-53735
15	38	49.4	469	6	US-10-505-928-457
16	38	49.4	483	7	US-11-105-233-180
17	38	49.4	1614	6	US-10-505-928-199
18	37	48.1	93	6	US-10-449-902-33339
19	37	48.1	304	6	US-10-449-902-24282
20	37	48.1	306	6	US-10-953-349-24281
21	37	48.1	322	6	US-10-953-349-24280
22	37	48.1	390	6	US-10-449-902-44395
23	37	48.1	520	7	US-11-293-697-3712
24	37	48.1	955	6	US-10-449-902-42940
25	36.5	47.4	515	6	US-10-953-349-21352

Sequence 21351, A  
Sequence 21350, A  
Sequence 31278, A  
Sequence 1326, Ap  
Sequence 396, App  
Sequence 33250, A  
Sequence 43998, A  
Sequence 56574, A  
Sequence 39097, A  
Sequence 39096, A  
Sequence 37830, A  
Sequence 51417, A  
Sequence 270, App  
Sequence 16, Appl  
Sequence 15077, A  
Sequence 12, Appl  
Sequence 27509, A  
Sequence 27508, A  
Sequence 220, App  
Sequence 1, Appl

#### ALIGNMENTS

##### RESULT 1

US-11-105-233-182  
; Sequence 182, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Thiagalingam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; FILE OF INVENTION: Tumors  
; FILE REFERENCE: 1657/2001  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; CURRENT FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 182  
; LENGTH: 400  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-105-233-182

Query Match 83.1%; Score 64; DB 7; Length 400;  
Best Local Similarity 86.7%; Pred. No. 0.001;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TMOELNSRLASYLDK 16  
Db 83 TMOELNSRLASYLDK 97

##### RESULT 2

US-11-105-233-181  
; Sequence 181, Application US/11105233  
; Publication No. US20060134653A1  
; GENERAL INFORMATION:  
; APPLICANT: Thiagalingam et al  
; TITLE OF INVENTION: Differential Expression of Genes in MSI  
; FILE OF INVENTION: Tumors  
; FILE REFERENCE: 1657/2001  
; CURRENT APPLICATION NUMBER: US/11/105,233  
; CURRENT FILING DATE: 2005-04-13  
; NUMBER OF SEQ ID NOS: 202  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 181  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-105-233-181

Query Match 79.2%; Score 61; DB 7; Length 430;

```
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 183
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-183

Query Match          71.4%; Score 55; DB 7; Length 424;
Best Local Similarity 78.6%; Pred. No. 0.04;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 MQELNSRLASYLDK 16
Db 74 MQNLNDRLASYLEK 87

RESULT 6
US-11-289-102-289
; Sequence 289, Application US/11289102
; Publication No. US20060121511A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Hyerim
; APPLICANT: Shaw, Peter M.
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO
; TITLE OF INVENTION: MICROTUBULE-STABILIZING AGENTS
; FILE REFERENCE: 10338 NP
; CURRENT APPLICATION NUMBER: US/11/289,102
; CURRENT FILING DATE: 2005-11-29
; PRIOR APPLICATION NUMBER: US 60/631,993
; PRIOR FILING DATE: 2004-11-30
; NUMBER OF SEQ ID NOS: 395
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 289
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-289-102-289

Query Match          67.5%; Score 52; DB 7; Length 466;
Best Local Similarity 64.3%; Pred. No. 0.15;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 MQELNSRLASYLDK 16
Db 107 LQELNDRFANYIDK 120

RESULT 7
US-11-270-040-4
; Sequence 4, Application US/11270040
; Publication No. US20060110762A1
; GENERAL INFORMATION:
; APPLICANT: Kapil, Sanjay
; APPLICANT: Kim, Jeong-Ki
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY
; TITLE OF INVENTION: SYNDROME VIRUS RECEPTOR COMPONENTS AND USES THEREOF
; FILE REFERENCE: 14337/004001
; CURRENT APPLICATION NUMBER: US/11/270,040
; CURRENT FILING DATE: 2005-11-09
; PRIOR APPLICATION NUMBER: 60/680,297
; PRIOR FILING DATE: 2005-05-12
; PRIOR APPLICATION NUMBER: 60/626,788
; PRIOR FILING DATE: 2004-11-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 551
; TYPE: PRT
```

---

```
Best Local Similarity 80.0%; Pred. No. 0.0037;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TMQELNSRLASYLDK 16
Db 83 TMQSLNDRLASLYLDR 97

RESULT 3
US-11-105-233-176
; Sequence 176, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; TITLE OF INVENTION: Tumors
; FILE REFERENCE: 1657/2001
; CURRENT APPLICATION NUMBER: US/11/105,233
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-105-233-176

Query Match          75.3%; Score 58; DB 7; Length 422;
Best Local Similarity 68.8%; Pred. No. 0.012;
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STMQELNSRLASYLDK 16
Db 74 ATMQNLNDRLASYVEK 89

RESULT 4
US-11-293-697-4615
; Sequence 4615, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4615
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-4615

Query Match          72.7%; Score 56; DB 7; Length 259;
Best Local Similarity 73.3%; Pred. No. 0.016;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TMQELNSRLASYLDK 16
Db 120 TMQSLNDRLASYLDR 134

RESULT 5
US-11-105-233-183
; Sequence 183, Application US/11105233
; Publication No. US20060134653A1
; GENERAL INFORMATION:
; APPLICANT: Thiagalingam et al
; TITLE OF INVENTION: Differential Expression of Genes in MSI
; TITLE OF INVENTION: Tumors
```



```
; ORGANISM: Cercopithecus aethiops
US-11-270-040-4
Query Match      67.5%; Score 52; DB 7; Length 551;
Best Local Similarity 64.3%; Pred. No. 0.18;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

QY 3 MOELNSRLASYLDK 16
Db 87 LQELNDRFANYIDK 100
:|||||:|:|:|

RESULT 8
US-10-505-928-314
; Sequence 314, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/33178
; CURRENT APPLICATION NUMBER: US/10/505,928
; PRIOR FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 314
; LENGTH: 586
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-505-928-314

Query Match      63.6%; Score 49; DB 6; Length 586;
Best Local Similarity 64.3%; Pred. No. 0.65;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

QY 3 MOELNSRLASYLDK 16
Db 36 LQELNDRFANYIDK 49
:|||||:|:|:|

RESULT 9
US-11-293-697-4090
; Sequence 4090, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; PRIOR FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4090
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-4090

Query Match      62.3%; Score 48; DB 7; Length 450;
Best Local Similarity 57.1%; Pred. No. 0.72;
Matches 8; Conservative 4; Mismatches 2; Indels 2; Gaps 0;

QY 3 MOELNSRLASYLDK 16
Db 92 LQELNDRFANYIEK 105
:|||||:|:|:|

RESULT 10
US-10-449-902-50320
; Sequence 50320, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50320
; LENGTH: 794
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-50320

Query Match      55.8%; Score 43; DB 6; Length 794;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 8; Conservative 4; Mismatches 4; Indels 4; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
Db 318 SAKWELSDNMSSYLDK 333
|:|:|:|:|:|:|

RESULT 11
US-10-449-902-47888
; Sequence 47888, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47888
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-47888

Query Match      55.8%; Score 43; DB 6; Length 969;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 8; Conservative 4; Mismatches 4; Indels 4; Gaps 0;

QY 1 STMQELNSRLASYLDK 16
Db 493 SAKWELSDNMSSYLDK 508
|:|:|:|:|:|:|

RESULT 12
US-11-293-697-4336
; Sequence 4336, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
```

```
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4336
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-4336

Query Match      53.2%; Score 41; DB 7; Length 535;
Best Local Similarity 53.8%; Pred. No. 15;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 QELNSRLASYLDK 16
Db 506 EEINFRLQYMDK 518
      :||: ||| :|||
      :||: ||| :|||

RESULT 13
US-11-293-697-3859
; Sequence 3859, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; PRIOR FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3859
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-3859

Query Match      49.4%; Score 38; DB 7; Length 328;
Best Local Similarity 88.9%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 SRLASYLDK 16
Db 299 SRLPSYLDK 307
      ||| |||||
      ||| |||||

RESULT 14
US-10-449-902-53735
; Sequence 53735, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53735
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Oryza sativa

US-10-449-902-53735
; Sequence 457, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 457
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-505-928-457

Query Match      49.4%; Score 38; DB 6; Length 469;
Best Local Similarity 42.9%; Pred. No. 42;
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 MOELNSRLASYLDK 16
Db 95 IKALNNKFASFDK 108
      :||: ||| :|||
      :||: ||| :|||

Search completed: July 5, 2006, 20:53:00
Job time : 9.16505 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2006, 19:18:42 ; Search time 36.0388 Seconds  
(without alignments)  
38.861 Million cell updates/sec

Title: US-10-774-928A-14  
Perfect score: 77  
Sequence: 1 STMQELNSRLASYLDK 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SID33/ptodata/2/iaa/5 COMB.pcp.\*
- 2: /EMC\_Celerra\_SID33/ptodata/2/iaa/6 COMB.pcp.\*
- 3: /EMC\_Celerra\_SID33/ptodata/2/iaa/7 COMB.pcp.\*
- 4: /EMC\_Celerra\_SID33/ptodata/2/iaa/H COMB.pcp.\*
- 5: /EMC\_Celerra\_SID33/ptodata/2/iaa/PCTUS COMB.pcp.\*
- 6: /EMC\_Celerra\_SID33/ptodata/2/iaa/RE COMB.pcp.\*
- 7: /EMC\_Celerra\_SID33/ptodata/2/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	100.0	16	2	US-09-809-003A-14
2	65	84.4	432	1	US-08-705-660-18
3	65	84.4	432	2	US-08-989-045-18
4	65	84.4	432	2	US-09-919-172-9
5	65	84.4	432	2	US-09-315-355A-18
6	64	83.1	22	1	US-07-934-656A-2
7	64	83.1	22	1	US-07-934-656A-3
8	64	83.1	22	1	US-07-934-656A-4
9	64	83.1	22	1	US-07-934-656A-6
10	64	83.1	22	1	US-07-934-656A-7
11	64	83.1	142	2	US-09-370-838-213
12	64	83.1	142	2	US-09-854-133-213
13	64	83.1	400	1	US-07-730-953-2
14	64	83.1	442	2	US-09-949-016-8625
15	64	83.1	456	2	US-09-919-172-31
16	64	83.1	471	2	US-09-538-092-837
17	64	83.1	593	2	US-09-538-092-919
18	61	79.2	120	2	US-09-621-976-5162
19	61	79.2	212	2	US-09-949-016-8463
20	61	79.2	422	1	US-09-067-351-3
21	61	79.2	422	2	US-09-360-490-3
22	61	79.2	430	2	US-09-919-039-105
23	61	79.2	447	2	US-09-949-016-11032
24	60	77.9	22	1	US-07-934-656A-5
25	60	77.9	355	2	US-10-104-047-2943
26	60	77.9	435	2	US-09-949-016-8415

27	60	77.9	435	2	US-09-949-016-8792	Sequence 8792, Ap
28	57	74.0	376	2	US-09-949-016-8106	Sequence 8106, Ap
29	57	74.0	377	2	US-09-949-016-8141	Sequence 8141, Ap
30	57	74.0	404	2	US-09-949-016-6202	Sequence 6202, Ap
31	57	74.0	416	2	US-09-949-016-6201	Sequence 6201, Ap
32	57	74.0	433	2	US-09-949-016-10900	Sequence 10900, A
33	57	74.0	436	2	US-09-949-016-6983	Sequence 6983, Ap
34	56	72.7	425	2	US-09-538-092-1366	Sequence 1366, Ap
35	54	70.1	247	2	US-10-104-047-3859	Sequence 3859, Ap
36	52	67.5	465	2	US-09-538-092-877	Sequence 877, App
37	52	67.5	466	2	US-09-610-401-3	Sequence 3, Appli
38	52	67.5	466	2	US-09-610-401-4	Sequence 4, Appli
39	52	67.5	466	2	US-09-167-206-12	Sequence 12, Appli
40	52	67.5	466	2	US-09-949-016-6351	Sequence 6351, Ap
41	52	67.5	466	3	US-10-152-647-3	Sequence 3, Appli
42	52	67.5	466	3	US-10-152-647-4	Sequence 4, Appli
43	52	67.5	506	2	US-09-949-016-7650	Sequence 7650, Ap
44	51	66.2	458	2	US-09-949-016-8791	Sequence 8791, Ap
45	50	64.9	120	2	US-09-370-838-201	Sequence 201, App

## ALIGNMENTS

## RESULT 1

US-09-809-003A-14  
; Sequence 14, Application US/09809003A  
; Patent No. 6673351

; GENERAL INFORMATION:  
; APPLICANT: O'Daly, Jose

; TITLE OF INVENTION: Compositions and Methods for the Treatment and Clinical Remission  
; TITLE OF INVENTION: Psoriasis

; FILE REFERENCE: 302178

; CURRENT APPLICATION NUMBER: US/09/809,003A

; CURRENT FILING DATE: 2001-03-16

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 14

; LENGTH: 16

; TYPE: PRT

; ORGANISM: Leishmania

US-09-809-003A-14

Query Match 100.0%; Score 77; DB 2; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STMQELNSRLASYLDK 16

Db 1 STMQELNSRLASYLDK 16

## RESULT 2

US-08-705-660-18

; Sequence 18, Application US/08705660

; Patent No. 5858683

; GENERAL INFORMATION:

; APPLICANT: KEESEE, SUSAN

; APPLICANT: OBAR, ROBERT

; APPLICANT: WU, YING-JYE

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE

; TITLE OF INVENTION: DETECTION OF CERVICAL CANCER

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Testa, Hurwitz & Thibault

; STREET: 125 High St.

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/705,660
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: GREENHALGH, DUNCAN A
; REGISTRATION NUMBER: 38,678
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-705-660-18

```

Query Match 84.4%; Score 65; DB 1; Length 432;  
 Best Local Similarity 81.2%; Pred. No. 0.0055;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 STMQELNSRLASYLDK 16
   :||| ||| ||| ||| |||
DB 86 ATMQLNDRSLASYLDK 101

```

```

RESULT 3
; Sequence 18, Application US/08989045
; Patent No. 6027905
; GENERAL INFORMATION:
; APPLICANT: KEESEE, SUSAN
; APPLICANT: OBAR, ROBERT
; APPLICANT: WU, YING-JYE
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; DETECTION OF CERVICAL CANCER
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,045
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GREENHALGH, DUNCAN A
; REGISTRATION NUMBER: 38,678
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-989-045-18

```

Query Match 84.4%; Score 65; DB 2; Length 432;  
 Best Local Similarity 81.2%; Pred. No. 0.0055;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 STMQELNSRLASYLDK 16
   :||| ||| ||| ||| |||
DB 86 ATMQLNDRSLASYLDK 101

```

```

RESULT 4
; Sequence 9, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673545 1454852CD1
; US-09-919-172-9

```

Query Match 84.4%; Score 65; DB 2; Length 432;  
 Best Local Similarity 81.2%; Pred. No. 0.0055;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 STMQELNSRLASYLDK 16
   :||| ||| ||| ||| |||
DB 86 ATMQLNDRSLASYLDK 101

```

```

RESULT 5
; Sequence 18, Application US/09315355A
; Patent No. 6803189
; GENERAL INFORMATION:
; APPLICANT: Keesee, Susan
; APPLICANT: Obar, Robert
; APPLICANT: Wu, Ying-Jye
; TITLE OF INVENTION: Methods for the Detection of Cervical Cancer
; FILE REFERENCE: MTP-023DV2
; CURRENT APPLICATION NUMBER: US/09/315,355A
; CURRENT FILING DATE: 1999-05-17
; PRIOR APPLICATION NUMBER: US 08/989,045
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: US 08/705,660
; PRIOR FILING DATE: 1996-08-30
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-315-355A-18

```

Query Match 84.4%; Score 65; DB 2; Length 432;  
 Best Local Similarity 81.2%; Pred. No. 0.0055;  
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 STMQELNSRLASYLDK 16
   :||| ||| ||| ||| |||

```

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
 STREET: 655 Fifteenth Street N.W. Suite 330  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005-5701  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/934,656A  
 FILING DATE: 27-JAN-1993  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DE P 40 23 945.4  
 FILING DATE: 27-JUL-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murray, Robert B.  
 REGISTRATION NUMBER: 22,980  
 REFERENCE/DOCKET NUMBER: P564-3003  
 TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (202)638-5000  
/ TELEFAX: (202)638-4810  
/ INFORMATION FOR SEQ ID NO: 4:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 22 amino acids  
/ TYPE: amino acid  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
US-07-934-656A-4

Query Match 83.1%; Score 64; DB 1; Length 22;  
Best Local Similarity 86.7%; Pred. No. 0.00034;  
Matches 13; Conservative 0; Mismatches 2; Indels 0;

QY 2 TMOELNSRLASYLDK 16  
Db |||||  
7 TMOELNSRLASYLDK 21

## RESULT 9

US-07-934-656A-5  
/ Sequence 6, Application US/07934656A  
/ Patent No. 5500347  
/ GENERAL INFORMATION:  
/ APPLICANT: MOLL, Roland  
/ APPLICANT: FRANK, Werner W.  
/ TITLE OF INVENTION: PROCESS FOR THE PURIFICATION OF  
/ CYCLOKATIN 20 AND ITS USE FOR THE PRODUCTION OF  
/ ANTIBODIES  
/ NUMBER OF SEQUENCES: 30  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Nikolaide, Marmelstein, Murray & Oram  
/ STREET: 655 Fifteenth Street N.W. Suite 330  
/ CITY: Washington  
/ STATE: D.C.  
/ COUNTRY: U.S.A.  
/ ZIP: 20005-5701  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patent In Release #1.0, Version #1.25  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/07/934, 656A  
/ FILING DATE: 27-JAN-1993  
/ CLASSIFICATION: 530  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: DE P 40 23 945.4  
/ FILING DATE: 27-JUL-1990  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Murray, Robert B.  
/ REGISTRATION NUMBER: 22,980  
/ REFERENCE/DOCKET NUMBER: P564-3003  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (202)638-5000  
/ TELEFAX: (202)638-4810  
/ INFORMATION FOR SEQ ID NO: 6:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 22 amino acids  
/ TYPE: amino acid  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
US-07-934-656A-6

Query Match 83.1%; Score 64; DB 1; Length 22;  
Best Local Similarity 86.7%; Pred. No. 0.00034;  
Matches 13; Conservative 0; Mismatches 2; Indels 0;

QY 2 TMOELNSRLASYLDK 16  
Db |||||  
7 TMOELNSRLASYLDK 21

## RESULT 10

US-07-934-656A-7  
/ Sequence 7, Application US/07934656A  
/ Patent No. 5500347  
/ GENERAL INFORMATION:  
/ APPLICANT: MOLL, Roland  
/ APPLICANT: FRANK, Werner W.  
/ TITLE OF INVENTION: PROCESS FOR THE PURIFICATION OF  
/ CYCLOKATIN 20 AND ITS USE FOR THE PRODUCTION OF  
/ ANTIBODIES  
/ NUMBER OF SEQUENCES: 30  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Nikolaide, Marmelstein, Murray & Oram  
/ STREET: 655 Fifteenth Street N.W. Suite 330  
/ CITY: Washington  
/ STATE: D.C.  
/ COUNTRY: U.S.A.  
/ ZIP: 20005-5701  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patent In Release #1.0, Version #1.25  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/07/934, 656A  
/ FILING DATE: 27-JAN-1993  
/ CLASSIFICATION: 530  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: DE P 40 23 945.4  
/ FILING DATE: 27-JUL-1990  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Murray, Robert B.  
/ REGISTRATION NUMBER: 22,980  
/ REFERENCE/DOCKET NUMBER: P564-3003  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (202)638-5000  
/ TELEFAX: (202)638-4810  
/ INFORMATION FOR SEQ ID NO: 7:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 22 amino acids  
/ TYPE: amino acid  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
US-07-934-656A-7

Query Match 83.1%; Score 64; DB 1; Length 22;  
Best Local Similarity 86.7%; Pred. No. 0.00034;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16  
Db |||||  
7 TMOELNSRLASYLDK 21

## RESULT 11

US-09-370-838-213  
/ Sequence 213, Application US/09370838  
/ Patent No. 644425  
/ GENERAL INFORMATION:  
/ APPLICANT: Reed, Steven G.  
/ APPLICANT: Lodes, Michael J.  
/ APPLICANT: Mohamath, Roadoh  
/ APPLICANT: Secrist, Heather  
/ TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
/ LUNG CANCER AND METHODS FOR THEIR USE  
/ FILE REFERENCE: 210121.475C1  
/ CURRENT APPLICATION NUMBER: US/09/370,838  
/ CURRENT FILING DATE: 1999-08-09  
/ EARLIER APPLICATION NUMBER: US 09/285,323  
/ EARLIER FILING DATE: 1999-04-02  
/ NUMBER OF SEQ ID NOS: 289  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 213

QY 2 TMOELNSRLASYLDK 16  
Db |||||  
7 TMOELNSRLASYLDK 21

```
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-213

  Query Match      83.1%; Score 64; DB 2; Length 142;
  Best Local Similarity 86.7%; Pred. No. 0.0025;
  Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16
   ||| ||| ||| ||| |||
Db 25 TMOQLNDRLASYLDK 39

RESULT 12
US-09-854-133-213
; Sequence 213, Application US/09854133
; Patent No. 6759508
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-213

  Query Match      83.1%; Score 64; DB 2; Length 142;
  Best Local Similarity 86.7%; Pred. No. 0.0025;
  Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16
   ||| ||| ||| ||| |||
Db 25 TMOQLNDRLASYLDK 39

RESULT 13
US-07-730-953-2
; Sequence 2, Application US/07730953
; Patent No. 5288614
; GENERAL INFORMATION:
; APPLICANT: BODENMULLER, Heinz
; APPLICANT: DESSAUER, Andreas
; TITLE OF INVENTION: METHOD FOR THE DETECTION OF MALIGNANT
; TITLE OF INVENTION: DISEASES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikolaide, Marmelstein, Murray & Oram
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/730,953
; FILING DATE: 19910723
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28

  Query Match      83.1%; Score 64; DB 2; Length 442;
  Best Local Similarity 86.7%; Pred. No. 0.0084;
  Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16
   ||| ||| ||| ||| |||
Db 125 TMOQLNDRLASYLDK 139

RESULT 15
US-09-919-172-31
; Sequence 31, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28

  Query Match      83.1%; Score 64; DB 2; Length 442;
  Best Local Similarity 86.7%; Pred. No. 0.0084;
  Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16
   ||| ||| ||| ||| |||
Db 125 TMOQLNDRLASYLDK 139

US-07-730-953-2
; Query Match      83.1%; Score 64; DB 1; Length 400;
; Best Local Similarity 86.7%; Pred. No. 0.0075;
; Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16
   ||| ||| ||| ||| |||
Db 83 TMOQLNDRLASYLDK 97

RESULT 14
US-09-949-016-8625
; Sequence 8625, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8625
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8625

  Query Match      83.1%; Score 64; DB 2; Length 442;
  Best Local Similarity 86.7%; Pred. No. 0.0084;
  Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TMOELNSRLASYLDK 16
   ||| ||| ||| ||| |||
Db 125 TMOQLNDRLASYLDK 139

US-09-919-172-31
; Sequence 31, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
```

```
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673545 1709387CD1
US-09-919-172-31
```

```
Query Match      83.1%; Score 64; DB 2: Length 456;
Best Local Similarity 86.7%; Pred. No. 0.0087;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      2 TMQELNSRLASYLDK 16
      ||| ||| ||| ||| |||
Db      108 TMQNLNDRLASLYLDK 122
```

```
Search completed: July 5, 2006, 19:24:23
Job time : 37.0388 secs
```